An Introduction to Differentiable Manifolds and Riemannian Geometry

Second Edition

William M. Boothby
DEPARTMENT OF MATHEMATICS
WASHINGTON UNIVERSITY
ST. LOUIS, MISSOURI

1986



ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Orlando San Diego New York Austin
London Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Boothby, William M. (William Munger), Date
An introduction to differentiable manifolds and Riemannian geometry.

(Pure and applied mathematics;)
Bibliography: p.
Includes index.
1. Differentiable manifolds. 2. Riemannian manifolds. I. Title. II. Series: Pure and applied mathematics (Academic Press) ;
QA3.P8 [QA614.3] 510 s [516.3'6] 85-13327
ISBN 0–12–116052–1 (alk. paper)
ISBN 0–12–116053–X (paperback)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

This book is dedicated with love and affection
to my wife, Ruth,
and to our sons,
Daniel, Thomas, and Mark.

Contents

IV. Vector Fields on a Manifold

V. Tensors and Tensor Fields on Manifolds

VI. Integration on Manifolds

VII. Differentiation on Riemannian Manifolds

VIII. Curvature

Preface to the Second Edition

This book has now been in print for ten years and has seen several printings. During this period it has been very gratifying to me to have heard from a surprisingly varied group of readers that it has been helpful to them in learning the fundamentals of differentiable manifolds and their geometry. It appears that to some extent at least the objectives set out in the original preface have been realized. Satisfying though this may be, ten years is more than enough time to accumulate a long list of annoying misprints and errors and a collection of possible improvements, many of which were suggested by users and readers. This makes it a real pleasure to be given a chance to rethink a number of proofs, add new problems, and clean up as many errors and misprints as could be found. There are, in fact, changes on almost every second or third page, some slight and some substantial: a number of proofs have been modified or improved; the historical notes have been enlarged; and the references have been updated. It is hoped that this has improved the readability while preserving whatever good qualities were there before.

I am very appreciative of the detailed comments and errata furnished by Chung-Shing Chen, H. V. Fagundes, J. F. C. Velson, and A. H. Clifford. I am profoundly grateful to the many students, colleagues, and friends encountered in many places, who have said kind and encouraging things, and to those who have taken the trouble to write to me about this book. This is a reward for my efforts which is all the more gratifying for having been unanticipated. It is hoped that readers of the first edition will find the second edition tidied up and improved and that new readers will find it helpful to them in learning an exciting and important field of mathematics.

I also wish to express my gratitude to the National Science Foundation for support at various times during which I have worked on this book.

Lavenen bei Gstaad
Switzerland

William M. Boothby

Preface to the First Edition

Apart from its own intrinsic interest, a knowledge of differentiable manifolds has become useful—even mandatory—in an ever-increasing number of areas of mathematics and of its applications. This is not too surprising, since differentiable manifolds are the underlying, if unacknowledged, objects of study in much of advanced calculus and analysis. Indeed, such topics as line and surface integrals, divergence and curl of vector fields, and Stokes's and Green's theorems find their most natural setting in manifold theory. But however natural the leap from calculus on domains of Euclidean space to calculus on manifolds may be to those who have made it, it is not at all easy for most students. It usually involves many weeks of concentrated work with very general concepts (whose importance is not clear until later) during which the relation to the already familiar ideas in calculus and linear algebra becomes lost—not irretrievably, but for all too long. Simple but nontrivial examples that illustrate the necessity for the high level of abstraction are not easy to present at this stage, and a realization of the power and utility of the methods must often be postponed for a dismayingly long time.

This book was planned and written as a text for a two-semester course designed, it is hoped, to overcome, or at least to minimize, some of these difficulties. It has, in fact, been used successfully several times in preliminary form as class notes for a two-semester course intended to lead the student from a reasonable mastery of advanced (multivariable) calculus and a rudimentary knowledge of general topology and linear algebra to a solid fundamental knowledge of differentiable manifolds, including some facility in working with the basic tools of manifold theory: tensors, differential forms, Lie and covariant derivatives, multiple integrals, and so on. Although in overall content this book necessarily overlaps the several available excellent books on manifold theory, there are differences in presentation and emphasis which, it is hoped, will make it particularly suitable as an introductory text.

To begin with, it is more elementary and less encyclopedic than most books on this subject. Special care has been taken to review, and then to develop, the connections with advanced calculus. In particular, all of Chapter II is devoted to functions and mappings on open subsets of Euclidean space, including a careful exposition and proof of the inverse function theorem. Efforts are made throughout to introduce new ideas gradually and with as much attention to intuition as possible. This has led to a longer but more readable presentation of inherently difficult material. When manifolds are first defined, an effort is made to have as many nontrivial examples as possible; for this reason, Lie groups, especially matrix groups, and certain quotient manifolds are introduced early and used throughout as examples. A fairly large number of problems (almost 400) is included to develop intuition and computational skills.

Further, it may be said that there has been a conscious effort to avoid or at least to economize generality insofar as that is possible. Concepts are often introduced in a rather ad hoc way with only the generality needed and, if possible, only when they are actually needed for some specific purpose. This is particularly noticeable in the treatment of tensors—which is far from general—and in the brief introduction to vector bundles (more specifically to the tangent bundle). Thus it is not claimed that this is a comprehensive book; the student will emerge with gaps in his knowledge of various subjects treated (for example, Lie groups or Riemannian geometry). On the other hand it is hoped that he will acquire strong motivation, computational skills, and a feeling for the subject that will make it easy for him to proceed to more advanced work in any of a number of areas using manifold theory: differential topology, Lie groups, symmetric and homogeneous spaces, harmonic analysis, dynamical systems, Morse theory, Riemann surfaces, and so on.

Finally, it should be said that the author has tried to include at every stage results that illustrate the power of these ideas. Chapter VI is especially noteworthy in this respect in that it includes complete proofs of Brouwer's fixed point theorem and of the nonexistence of nowhere-vanishing continuous vector fields on even-dimensional spheres. In a similar vein, the existence of a bi-invariant measure on compact Lie groups is demonstrated and applied to prove the complete reducibility of their linear representations. Then, in a later chapter, compact groups are used as simple examples of symmetric spaces, and their corresponding geometry is used to prove that every element lies on a one-parameter subgroup. In the last two chapters, which deal with Riemannian geometry of abstract n-dimensional manifolds, the relation to the more easily visualized geometry of curves and surfaces in Euclidean space is carefully spelled out and is used to develop the general ideas for which such applications as the Hopf–Rinow theorem are given. Thus, by a selection of accessible but important applications, some truly

nontrivial, unexpected (to the student) results are obtained from the abstract machinery so patiently constructed.

Briefly, the organization of the book is as follows. Chapter I is a very intuitive introduction and fixes some of the conventions and notation that are used. Chapter II is largely advanced calculus and may very well be omitted or skimmed by better prepared readers. In Chapter III, the basic concept of differentiable manifold is introduced along with mappings of manifolds and their properties; a fairly extensive discussion of examples is included. Chapter IV is particularly concerned with vectors and vector fields and with a careful exposition of the existence theorem for solutions of systems of ordinary differential equations and the related one-parameter group action. In Chapter V covariant tensors and differential forms are treated in some detail and then used to develop a theory of integration on manifolds in Chapter VI. Numerous applications are given. It would be possible to use Chapters II–VI as the basis of a one-semester course for students who wish to learn the fundamentals of differentiable manifolds without any Riemannian geometry. On the other hand, for students who already have some experience with manifolds, Chapters VII and VIII could serve as a brief introduction to Riemannian geometry. In these last two chapters, beginning from curves and surfaces in Euclidean space, the concept of Riemannian connection and covariant differentiation is carefully developed and used to give a fairly extensive discussion of geodesics—including the Hopf–Rinow theorem—and a shorter treatment of curvature. The natural (bi-invariant) geometry on compact Lie groups and Riemannian manifolds of constant curvature are both discussed in some detail as examples of the general theory. This discussion is based on a fairly complete treatment of covering spaces, discontinuous group action, and the fundamental group given earlier in the book.

This book, as do many of the books in this subject, owes much to the influence of S. S. Chern. For many years his University of Chicago notes——still an important reference (Chern [1])—were virtually the only systematic account of most of the topics in this text. Even more importantly his courses, lectures, published works, and above all his personal encouragement have had an impressive influence on a whole generation of differential geometers, among whom this author had the good fortune to be included. Another source of inspiration to the author was the work of John Milnor. The manner in which he has made exciting fundamental research in differential topology and geometry available to specialist and nonspecialist alike through many careful expository works (written in a style that this author particularly admires) certainly deserves gratitude. No better material for further or supplemental reading to this text could be suggested than Milnor's two books [1] and [2].

For part of the time during which this book was being written, the

author benefitted from a visiting professorship at the University of Strasbourg, France, and he is particularly grateful for the opportunity to work there, in an atmosphere so conducive to advancing in the task he had undertaken.

The author would also like to acknowledge with gratitude the help given to him by his son, Thomas Boothby, by students and colleagues at Washington University, especially Humberto Alagia and Eduardo Cattani, and by Mrs. Virginia Hundley for her careful work preparing the manuscript.

I INTRODUCTION TO MANIFOLDS

In this chapter, we establish some preliminary notations and give an intuitive, geometric discussion of a number of examples of manifolds—the primary objects of study throughout the book. Most of these examples are surfaces in Euclidean space; for these—together with curves on the plane and in space—were the original objects of study in classical differential geometry and are the source of much of the current theory.

The first two sections deal primarily with notational matters and the relation between Euclidean space, its model $\boldsymbol{R}^n$, and real vector spaces. In Section 3 a precise definition of topological manifolds is given, and in the remaining sections this concept is illustrated.

1 Preliminary Comments on $\boldsymbol{R}^n$

Let $\boldsymbol{R}$ denote the real numbers and $\boldsymbol{R}^n$ their n-fold Cartesian product

$$\overbrace{\boldsymbol{R} \times \cdots \times \boldsymbol{R}}^{n},$$

the set of all ordered n-tuples $(x^1, \ldots, x^n)$ of real numbers. Individual n-tuples may be denoted at times by a single letter. Thus $x = (x^1, \ldots, x^n)$, $a = (a^1, \ldots, a^n)$, and so on. We agree once and for all to use on $\boldsymbol{R}^n$ its topology as a metric space with the metric defined by

$$d(x, y) = \left(\sum_{i=1}^{n} (x^i - y^i)^2 \right)^{1/2}.$$

The neighborhoods are then open balls $B_\varepsilon^n(x)$, or $B_\varepsilon(x)$ or, equivalently, open cubes $C_\varepsilon^n(x)$, or $C_\varepsilon(x)$ defined for any $\varepsilon > 0$ as

$$B_\varepsilon(x) = \{y \in \boldsymbol{R}^n \mid d(x, y) < \varepsilon\},$$

and

$$C_\varepsilon(x) = \{y \in \boldsymbol{R}^n \mid |x^i - y^i| < \varepsilon, i = 1, \dots, n\},$$

a cube of side 2ε and center x. Note that $\boldsymbol{R}^1 = \boldsymbol{R}$ and we define $\boldsymbol{R}^0$ to be a single point.

We shall invariably consider $\boldsymbol{R}^n$ with the topology defined by the metric. This space $\boldsymbol{R}^n$ is used in several senses, however, and we must usually decide from the context which one is intended. Sometimes $\boldsymbol{R}^n$ means merely $\boldsymbol{R}^n$ as *topological* space, sometimes $\boldsymbol{R}^n$ denotes an n-dimensional vector space, and sometimes it is identified with Euclidean space. We will comment on this last identification in Section 2 and examine here the other meanings of $\boldsymbol{R}^n$.

We assume that the definition and basic theorems of vector spaces over $\boldsymbol{R}$ are known to the reader. Among these is the theorem which states that any two vector spaces over $\boldsymbol{R}$ which have the same dimension n are isomorphic. It is important to note that this isomorphism depends on *choices* of *bases* in the two spaces; there is in general no *natural* or *canonical* isomorphism independent of these choices. However, there does exist one important example of an n-dimensional vector space over $\boldsymbol{R}$ which has a distinguished or canonical basis—a basis which is somehow given by the nature of the space itself. We refer to the vector space of n-tuples of real numbers with componentwise addition and scalar multiplication. This is, as a set at least, just $\boldsymbol{R}^n$; should we wish on occasion to avoid confusion, then we will denote it by V^n (and use boldface for its elements ($\mathbf{x}$ instead of x, and so forth). For this space the n-tuples $\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, 0, \dots, 0, 1)$ are a basis, known as the *natural* or *canonical* basis. We may at times suppose that the n-tuples are written as rows, that is, $1 \times n$ matrices, and at other times as columns, that is, $n \times 1$ matrices. This only becomes important should we wish to use matrix notation to simplify things a bit, for example, to describe linear mappings, equations, and so on.

Thus $\boldsymbol{R}^n$ *may* denote a vector space of dimension n over $\boldsymbol{R}$. We sometimes mean even more by $\boldsymbol{R}^n$. An abstract n-dimensional vector space over $\boldsymbol{R}$ is called *Euclidean* if it has defined on it a positive definite inner product. In general there is no natural way to choose such an inner product, but in the case of $\boldsymbol{R}^n$ or V^n, again we have the natural inner product

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x^i y^i.$$

It is characterized by the fact that relative to this inner product the natural basis is orthonormal, $(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$.

Thus at times $\boldsymbol{R}^n$ is a Euclidean vector space, but one which has a built-in orthonormal basis and inner product. An abstract vector space, even if Euclidean, does not have any such preferred basis. The metric in $\boldsymbol{R}^n$ discussed at the beginning can be defined using the inner product on $\boldsymbol{R}^n$. We define $\|\mathbf{x}\|$, the *norm* of the vector $\mathbf{x}$, by $\|\mathbf{x}\| = ((\mathbf{x}, \mathbf{x}))^{1/2}$. Then we have

$$d(x, y) = \|\mathbf{x} - \mathbf{y}\|.$$

This notation is frequently useful even when we are dealing with $\boldsymbol{R}^n$ as a metric space and not using its vector space structure. Note, in particular, that $\|\mathbf{x}\| = d(x, 0)$, the distance from the point x to the origin. In this equality $\mathbf{x}$ is a vector on the left-hand side, and x is the corresponding point on the right-hand side; an illustration of the way various interpretations of $\boldsymbol{R}^n$ can be mixed together.

Exercises

1. Show that if A is an $m \times n$ matrix, then the mapping from V^n to V^m (with elements written as $n \times 1$ and $m \times 1$ matrices), which is defined by $\mathbf{y} = A\mathbf{x}$, is continuous. Identify the images of the canonical basis of V^n as linear combinations of the canonical basis of V^m.
2. Find conditions for the mapping of Exercise 1 to be onto; to be one-to-one.
3. Show that if W is an n-dimensional Euclidean vector space, then there exists an isometry, that is, an isomorphism preserving the inner product, of W onto $\boldsymbol{R}^n$ interpreted as Euclidean vector space.
4. Show that $\boldsymbol{C}^n$, the space of n-tuples of complex numbers, may be placed in one-to-one correspondence with $\boldsymbol{R}^{2n}$. Can this correspondence be a vector space isomorphism?
5. Exhibit an isomorphism between the vector space of $m \times n$ matrices over $\boldsymbol{R}$ and the vector space $\boldsymbol{R}^{mn}$. Show that the map $X \to AX$, where A is a fixed $m \times m$ matrix and X is an arbitrary $m \times n$ matrix (over $\boldsymbol{R}$), is continuous in the topology derived from $\boldsymbol{R}^{mn}$.
6. Show that $\|\mathbf{x}\|$ has the following properties:

   (a) $\|\mathbf{x} \pm \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$;
   (b) $\|\mathbf{x}\| - \|\mathbf{y}\| \le \|\mathbf{x} - \mathbf{y}\|$;
   (c) $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|, \alpha \in \boldsymbol{R}$; and
   (d) explain how (a) is related to the triangle inequality of $d(x, y)$.

7. Show that an isometry of a Euclidean vector space onto itself has an orthogonal matrix relative to any orthonormal basis.
8. Prove that every Euclidean vector space V has an orthonormal basis. Construct your proof in such a way that if W is a given subspace of V, $\dim W = r$, then the first r vectors of the basis of V are a basis of W.

2 $\boldsymbol{R}^n$ and Euclidean Space

Another role which $\boldsymbol{R}^n$ plays is that of a model for n-dimensional *Euclidean* space $\boldsymbol{E}^n$, in the sense of Euclidean geometry. In fact many texts simply refer to $\boldsymbol{R}^n$ with the metric $d(x, y)$ as Euclidean space. This identification is misleading in the same sense that it would be misleading to identify *all* n-dimensional vector spaces with $\boldsymbol{R}^n$; moreover unless clearly understood, it is an identification that can hamper clarification of the concept of manifold and the role of coordinates. Certainly Euclid and the geometers before the seventeenth century did not think of the Euclidean plane or three-dimensional space—which we denote by $\boldsymbol{E}^2$ and $\boldsymbol{E}^3$—as pairs or triples of real numbers. In fact they were defined axiomatically beginning with undefined objects such as points and lines together with a list of their properties—the axioms—from which the theorems of geometry were then deduced. This is the path which we all follow in learning the basic ideas of Euclidean plane and solid geometry, about which most of us know quite a bit before studying analytic or coordinate geometry at all. The identification of $\boldsymbol{R}^n$ and $\boldsymbol{E}^n$ came about after the invention of analytic geometry by Fermat and Descartes and was eagerly seized upon since it is very tricky and difficult to give a suitable definition of Euclidean space, of any dimension, in the spirit of Euclid, that is, by giving axioms for (abstract) Euclidean space as one does for abstract vector spaces. This difficulty was certainly recognized for a very long time, and has interested many great mathematicians. It led in part to the discovery of non-Euclidean geometries and thus to manifolds. A careful axiomatic definition of Euclidean space is given by Hilbert [1]. Since our use of Euclidean geometry is mainly to aid our intuition, we shall be content with assuming that the reader "knows" this geometry from high school.

Consider the Euclidean plane $\boldsymbol{E}^2$ as studied in high school geometry; definitions are made, theorems proved, and so on, *without* coordinates. One later introduces coordinates using the notions of length and perpendicularity in choosing two mutually perpendicular "number axes" which are used to define a one-to-one mapping of $\boldsymbol{E}^2$ onto $\boldsymbol{R}^2$ by $p \to (x(p), y(p))$, the coordinates of $p \in \boldsymbol{E}^2$. This mapping is (by design) an isometry, preserving distances of points of $\boldsymbol{E}^2$ and their images in $\boldsymbol{R}^2$. Finally one obtains further correspondences of essential geometric elements, for example, lines of $\boldsymbol{E}^2$ with subsets of $\boldsymbol{R}^2$ consisting of the solutions of linear equations. Thus we carry each geometric object to a corresponding one in $\boldsymbol{R}^2$. It is the existence of such coordinate mappings which make the identification of $\boldsymbol{E}^2$ and $\boldsymbol{R}^2$ possible. But caution! An *arbitrary choice* of coordinates is involved, there is no *natural, geometrically determined* way to identify the two spaces. Thus, at best, we can say that $\boldsymbol{R}^2$ may be identified with $\boldsymbol{E}^2$ *plus a coordinate system.* Even then we need to define in $\boldsymbol{R}^2$ the notions of line, angle of lines, and

other attributes of the Euclidean plane before thinking of it as Euclidean space. Thus, with qualifications, we may identify $\boldsymbol{E}^2$ and $\boldsymbol{R}^2$ or $\boldsymbol{E}^n$ and $\boldsymbol{R}^n$, especially remembering that they carry a choice of rectangular coordinates.

We conclude with a brief indication of why we might not always wish to make the identification, that is, to use the analytic geometry approach to the study of a geometry. Whenever $\boldsymbol{E}^n$ and $\boldsymbol{R}^n$ are identified it involves the choice of a coordinate system, as we have seen. It then becomes difficult at times to distinguish underlying geometric properties from those which depend on the choice of coordinates. An example: Having identified $\boldsymbol{E}^2$ and $\boldsymbol{R}^2$ and lines with the graphs of linear equations, for instance,

$$L = \{(x, y) \mid y = mx + b\},$$

we define the *slope m* and the *y-intercept b*. Neither has geometric meaning; they depend on the choice of coordinates. However, given two such lines of slope m_1, m_2, the expression $(m_2 - m_1)/(1 + m_1 m_2)$ does have geometric meaning. This can be demonstrated by directly checking independence of the choice of coordinates—a tedious process—or determining that its value is the tangent of the angle between the lines, a concept which is independent of coordinates! It should be clear that it can be difficult to do geometry, even in the simplest case of Euclidean geometry, working with coordinates alone, that is, with the model $\boldsymbol{R}^n$. We need to develop both the coordinate method and the coordinate-free method of approach. Thus we shall often seek ways of looking at manifolds and their geometry which do not involve coordinates, but will use coordinates as a useful computational device (and more) when necessary.

However, being aware now of what is involved, we shall usually refer to $\boldsymbol{R}^n$ as Euclidean space and make the identification. This is especially true when we are interested only in questions involving topology—as in the next section—or differentiability.

Exercises

1. Using standard equations for change of Cartesian coordinates, verify that $(m_2 - m_1)/(1 + m_1 m_2)$ is independent of the choice of coordinates.
2. Similarly, show that $((x_2 - x_1)^2 + (y_2 - y_1)^2)^{1/2}$ is a function of points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ which does not depend on the choice of coordinates.
3. How do we describe the subset of $\boldsymbol{R}^n$ which corresponds to a segment $\overline{pq}$ in $\boldsymbol{E}^n$? to a line? to a 2-plane not through the origin?

If we wish to prove the theorems of Euclidean geometry by analytical geometry methods, we need to define the notion of congruence. We say that two figures are *congruent* if there is a *rigid motion* of the space, that is, an isometry or distance-preserving transformation, which carries one figure to the other.

4. Identifying E^2 with R^2, describe analytically the rigid motions of R^2. Show that they form a group.
5. Using Exercise 4 prove that two triangles are congruent if and only if corresponding sides are of equal length.

3 Topological Manifolds

Of all the spaces which one studies in topology the Euclidean spaces and their subspaces are the most important. As we have just seen, the metric spaces R^n serve as a *topological model* for Euclidean space E^n, for finite-dimensional vector spaces over R or C, and for other basic mathematical systems which we shall encounter later. It is natural enough that we are led to study those spaces which are *locally* like R^n, more precisely those spaces for which each point p has a neighborhood U which is homeomorphic to an open subset U' of R^n, n fixed. We say that a space with this property is *locally Euclidean of dimension n*, and in order to stay as close as possible to Euclidean spaces, we will consider spaces called *manifolds*, defined as follows.

(3.1) Definition *A manifold M* of *dimension n*, or *n-manifold*, is a topological space with the following properties:

(i) M is Hausdorff,
(ii) M is locally Euclidean of dimension n, and
(iii) M has a countable basis of open sets.

As a matter of notation dim M is used for the *dimension* of M; when dim $M = 0$, then M is a countable space with the discrete topology. It follows from the homeomorphism of U and U' that *locally Euclidean* is equivalent to the requirement that each point p have a neighborhood U homeomorphic to an n-ball in R^n. Thus a manifold of dimension 1 is locally homeomorphic to an open interval, a manifold of dimension 2 is locally homeomorphic to an open disk, and so on. Our first examples will remove any lingering suspicion that an n-manifold is necessarily globally equivalent, that is, homeomorphic, to E^n.

(3.2) Example Let M be an open subset of R^n with the subspace topology; then M is an n-manifold.

Indeed properties (i) and (iii) of Definition 3.1 are hereditary, holding for any subspace of a space which possesses them; and we see that (ii) holds with $U = U' = M$ and with the homeomorphism of U to U' being the identity map. A bit of imagination, assisted perhaps by Fig. I.1, will show that even

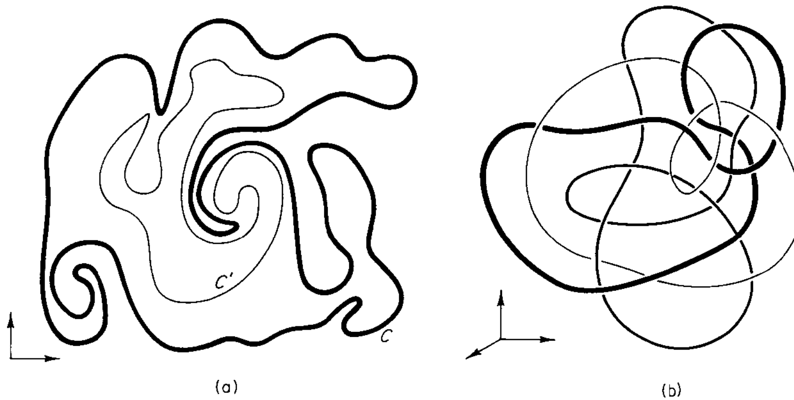


Figure I.1

(a) The manifold is the open set M of $\boldsymbol{R}^2$ between the curves C and C'. (b) The manifold is the open subset of $\boldsymbol{R}^3$ obtained by removing the knots.

when $n = 2$ or 3 these examples can be rather complicated and certainly not equivalent to Euclidean space in general, although they may be in special cases: a trivial such case is $M = E^n$.

(3.3) Example The simplest examples of manifolds not homeomorphic to open subsets of Euclidean space are the circle S^1 and the 2-sphere S^2, which may be defined to be all points of $\boldsymbol{E}^2$, or of $\boldsymbol{E}^3$, respectively, which are at unit distance from a fixed point 0.

These are to be taken with the subspace topology so that (i) and (iii) are immediate. To see that they are locally Euclidean we introduce coordinate axes with 0 as origin in the corresponding ambient Euclidean space. Thus in the case of S^2 we identify $\boldsymbol{R}^3$ and $\boldsymbol{E}^3$, and S^2 becomes the unit sphere centered at the origin. At each point p of S^2 we have a tangent plane and a unit normal vector N_p. There will be a coordinate axis which is *not* perpendicular to N_p and some neighborhood U of p on S^2 will then project in a continuous and one-to-one fashion onto an open set U' of the coordinate plane perpendicular to that axis. In Fig. I.2a, N_p is not perpendicular to the x_2-axis so for $q \in U$, the projection is given quite explicitly by $\varphi(q) = (x^1(q), 0, x^3(q))$, where $(x^1(q), x^2(q), x^3(q))$ are the coordinates of q in $\boldsymbol{E}^3$. Similar considerations show that S^1 is locally Euclidean. Note that S^2 and R^2 cannot be homeomorphic since one is compact while the other is not.

(3.4) Example Our final example is that of the surface of revolution obtained by revolving a circle around an axis which does not intersect it. The figure we obtain is the *torus* or "inner tube" (denoted T^2) as shown in Fig. I.2b. This figure can be studied analytically; it is easy to write down an

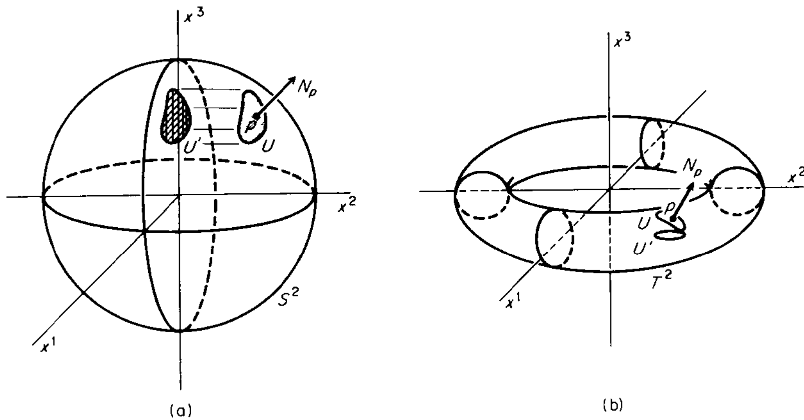


Figure I.2

(a) The spherical surface S^2 as a manifold. (b) The torus as a manifold.

equation whose locus it is if we introduce coordinates in E^3 as shown in the figure. In order to convince ourselves that it is indeed locally Euclidean we consider once more the normal vector N_p at $p \in T^2$. There will be at least one coordinate axis to which it is not perpendicular, say x^3. Then some neighborhood U of p projects homeomorphically onto a neighborhood U' in the x^1x^2-plane as illustrated. Since we use the relative topology derived from E^3, the space T^2 is necessarily Hausdorff and has a countable basis of open sets. Thus conditions (i)–(iii) of Definition 3.1 are satisfied.

(3.5) Remark It should be clear from the last two examples that certain subspaces M of E^3 are easily seen to be 2-manifolds; they are surfaces which are "smooth," that is, without corners or edges, so that they have at each $p \in M$ a (unit) normal vector N_p and tangent plane $T_p(M)$—to introduce notation we use later—which varies continuously as we move from point to point. (By this last requirement we mean that the components of the unit normal vector depend continuously on the point p.) This smoothness allows us to prove the locally Euclidean property by projection of a neighborhood of p onto a plane as in Examples 3.3 and 3.4. The other properties are immediate since we use the subspace topology. Figure I.3 shows some further examples of manifolds which can be obtained in this way. Obviously this method will not always work: The surface of a cube is a 2-manifold, in fact it is homeomorphic to S^2; but it has no tangent plane or normal vector at the corners and edges.

Example 3.2 gives an inkling at least, of how nasty a space can be and still be a manifold, even when it is connected—which we do not suppose in general. The following theorem will offer some reassurance.

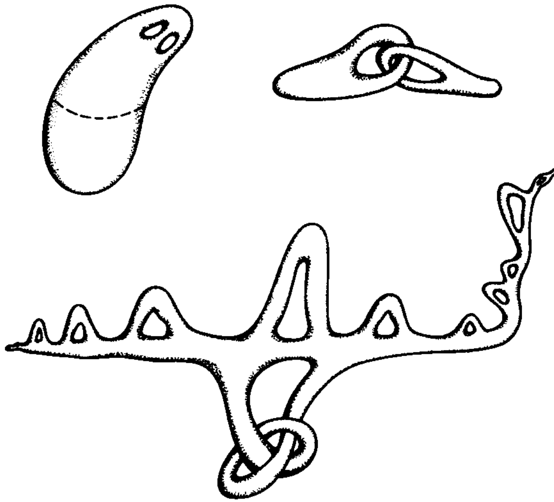


Figure I.3

(3.6) Theorem *A topological manifold M is locally connected, locally compact, and a union of a countable collection of compact subsets; furthermore, it is normal and metrizable.*

Proof These are all immediate consequences of the definition and standard theorems of general topology. Let p be a point of M and U a neighborhood of p homeomorphic to an open ball $B_\varepsilon(x)$ of radius ε in $\boldsymbol{R}^n$. We denote this homeomorphism by φ, and we suppose $\varphi(p) = x$. Then it is clear that interior to any neighborhood V of p there is a neighborhood W whose closure $\overline{W}$ is in U and for which $\varphi(W) = B_\delta(x)$ with $\varepsilon > \delta > 0$. It follows that M is locally connected at p since $B_\delta(x)$ and hence W, to which it is homeomorphic by φ^{-1}, is connected. Similarly $\overline{W}$ is compact since $\overline{B}_\delta(x)$ is compact; thus M is locally compact. Because M has a countable base of open sets, we may now suppose that it has a countable base of relatively compact open sets $\{V_i\}$; obviously $M = \bigcup \overline{V}_i$. Normality follows from Lindelöf's theorem and metrizability is then a consequence of the Urysohn metrization theorem (see Kelley [1]).

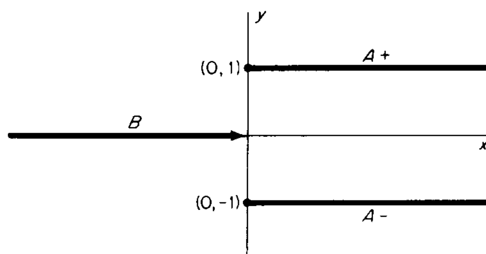
There is one difficulty in our concept of manifold about which we can do nothing at present. In fact it concerns Euclidean spaces and their topology. and arises even before consideration of manifolds: it is the question of dimension. Could it be that $\boldsymbol{E}^n$ and $\boldsymbol{E}^m$ are homeomorphic, or locally homeomorphic—so that an open set U of $\boldsymbol{E}^n$ is homeomorphic to some open set U' of $\boldsymbol{E}^m$ with $m \neq n$? The answer is no, but the proof is difficult and requires algebraic topology. It was proved in 1911 by L. E. J. Brouwer and is

known as Brouwer's theorem on invariance of domain. For a proof see Hurewicz and Wallman [1]. Later we shall be able to give a differentiable version of this theorem which will be sufficient for our needs; in this chapter we assume the theorem.

We make one final remark which connects this section with the preceding one. The notion of *coordinates* plays an important role in manifold theory, just as it does in the study of the geometry of $\boldsymbol{E}^n$. In $\boldsymbol{E}^n$, however, it is possible to find a single system of coordinates for the *entire* space, that is, to establish a correspondence between all of $\boldsymbol{E}^n$ and $\boldsymbol{R}^n$. Built into the definition of n-manifold M is a correspondence of a neighborhood U of each $p \in M$ and an open subset U' of $\boldsymbol{R}^n$. Letting $\varphi : U \to U'$ be this correspondence, we call the pair U, φ a *coordinate neighborhood* and the numbers $x^1(q), \ldots, x^n(q)$, given by $\varphi(q) = (x^1(q), \ldots, x^n(q))$, the *coordinates* of $q \in M$. We have assumed that this φ is a homeomorphism: it is one-to-one and φ and φ^{-1} are continuous. Thus each $q \in U$ has n uniquely determined coordinates, real numbers, which vary continuously with q. Of course the function $q \to x^i(q)$, which gives the ith coordinate, $1 \leq i \leq n$, is continuous; it is called the *ith coordinate function*. There is obviously nothing unique about our choice of coordinates; in Examples 3.3 and 3.4, we could equally well project the neighborhood of p discussed there to other coordinate planes. Finally note that even in the case of Euclidean space it is often useful to use *local* coordinates; the domain of a polar coordinate system on $\boldsymbol{E}^2$, for example, must omit a ray if it is to be one-to-one.

Exercises

1. Consider the following subset of $\boldsymbol{R}^2 : X = A_+ \cup A_- \cup B$ with
$A_+ = \{(x, y) \mid x \geq 0, y = +1\}$,
$A_- = \{(x, y) \mid x \geq 0, y = -1\}$,
$B = \{(x, y) \mid x < 0, y = 0\}$.



Define a topology as follows: We use the subspace topology (open intervals as a basis) on $A_+ - \{(0, 1)\}$, $A_- - \{(0, -1)\}$ and B; then for $\varepsilon > 0$ we let $N_\varepsilon^\pm = \{(x, \pm 1) \mid 0 \leq x < \varepsilon\} \cup \{(x, 0) \mid -\varepsilon \leq x < 0\}$ and use

N_ε^+ and N_ε^- as a basis of neighborhoods of (0, 1) and (0, −1), respectively. Show that the space X is locally Euclidean but is not a manifold.

A Hausdorff space M is said to be *paracompact* if every covering $\{U_\alpha\}$ of M by open sets has a *locally finite refinement*; more precisely, there is a covering $\{V_\beta\}$ which (i) *refines* $\{U_\alpha\}$ in the sense that each $V_\beta \subset U_\alpha$ for some α, and which (ii) is locally finite, that is, each $p \in M$ has a neighborhood W which intersects only a finite number of sets V_β.

2. Show that a manifold is paracompact. Show that a locally Euclidean, paracompact, Hausdorff space need not have a countable basis.
3. Show that a connected manifold M is pathwise connected, that is, $p, q \in M$ implies that there exists a continuous curve $f(s)$, $0 \le s \le 1$, with $f(0) = p, f(1) = q$.
4. Show that the (connected) components of a manifold M are open sets and are countable in number.

4 Further Examples of Manifolds. Cutting and Pasting

A hemispherical cap (including the equator) or a right circular cylinder (including the circles at the ends) are typical examples of manifolds with boundary. Except for the equator, or the end-circles, they are 2-manifolds and these boundary sets are themselves manifolds of dimension one less. In fact, they are homeomorphic to S^1 or to $S^1 \cup S^1$ in these two cases. An even simpler example is the upper half-plane H^2, or more generally H^n, where we shall mean by H^n the *subspace* of $\boldsymbol{R}^n$ defined by

$$H^n = \{(x^1, \ldots, x^n) \in \boldsymbol{R}^n \mid x^n \ge 0\}.$$

Every point $p \in H^n$ has a neighborhood U which is homeomorphic to an open subset U' of $\boldsymbol{R}^n$ *except* the set of points $(x^1, \ldots, x^{n-1}, 0)$, which obviously forms a subspace homeomorphic to $\boldsymbol{R}^{n-1}$, called the *boundary* of H^n and denoted by ∂H^n.

We shall define a *manifold with boundary* to be a Hausdorff space M with a countable basis of open sets which has the property that each $p \in M$ is contained in an open set U with a homeomorphism φ to either (a) an open set U' of $H^n - \partial H^n$ *or* (b) to an open set U' of H^n with $\varphi(p) \in \partial H^n$, that is, a boundary point of H^n. It can be shown (as a consequence of invariance of domain) that $p \in M$ is in one class or the other but not both; those p of the first type are called *interior points* of M and those p mapped onto the boundary of H^n by one, and hence by all, homeomorphisms of their neighborhoods into H^n are called *boundary points*. The collection of boundary points is then denoted by ∂M and is called the *boundary of M*. It is a manifold of dimension $n - 1$. We make no attempt to prove these facts here, but they will be discussed briefly in Chapter VI.

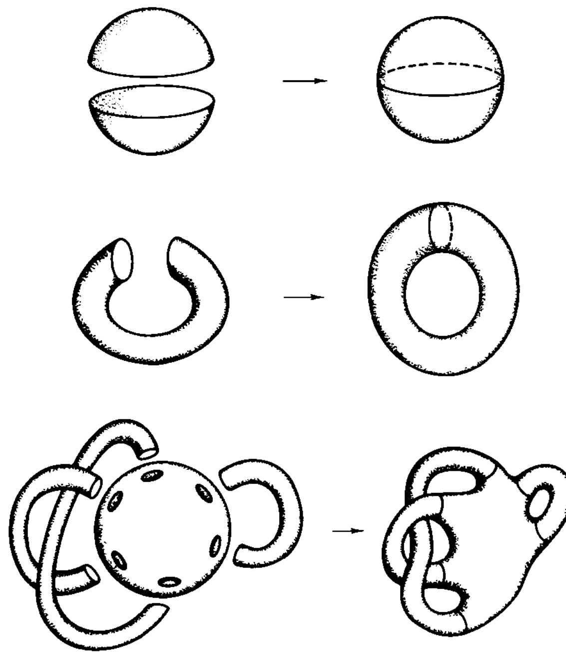


Figure I.4
Some examples of pasting.

Our interest is in pointing out that new surfaces, that is, 2-manifolds, can be formed by fastening together manifolds with boundary along their boundaries, that is, by identifying points of various boundary components by a homeomorphism, assuming of course the necessary condition that such components are homeomorphic. The simplest examples are S^2, which is obtained by pasting two disks (or hemispheres) together so as to form the equator, and T^2, formed by pasting the two end-circles of a cylinder together. However, one can go much further and paste any number of cylinders onto a sphere S^2 with "holes," that is, with circular disks removed. This gives various pretzellike surfaces as illustrated in Fig. I.4. We leave as an exercise the proof that these are manifolds. Thus to generate new 2-manifolds from old ones we may (1) cut out two disks, leaving a manifold M whose boundary ∂M is the disjoint union of two circles, and (2) paste on a cylinder or "handle" so that each end-circle is identified with one of the boundary circles of M.

The pasting on of handles is not the only way in which we can generate examples of 2-manifolds. It is also possible to do so by identifying or pasting together the edges of certain polygons. For example, the sides of a square may be identified in various ways in order to obtain surfaces. Figure I.5

illustrates this: we obtain a cylinder, Möbius band, torus, and Klein bottle. The latter cannot be pictured as a surface in $\boldsymbol{E}^3$ unless we allow it to cut itself as shown. Thus as a subspace of $\boldsymbol{E}^3$ it is *not* a manifold: it *is* possible to identify the sides of the square, as shown, and obtain a manifold—but it is *not* possible to put it inside $\boldsymbol{E}^3$.

For connected 2-manifolds M which lie smoothly inside $\boldsymbol{E}^3$ so that there is a tangent plane and normal line L_p at each point p, we may ask whether it is possible to choose a unit normal vector N_p (on L_p) for every $p \in M$ which varies continuously with M. It is easy to see that this is possible for S^2 and T^2 but not for the Möbius band (which is actually a manifold with boundary) or the Klein bottle. We say that a manifold or manifold with boundary is *orientable* if such a choice of N_p is possible. The following is a fundamental theorem of 2-manifolds.

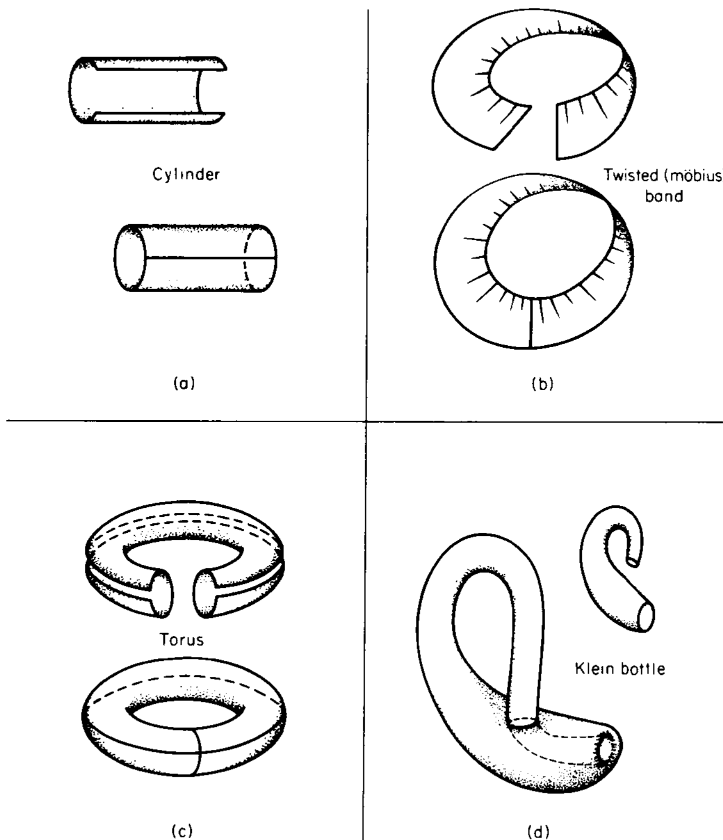


Figure I.5

Four ways to identify sides of a rectangle: (a) cylinder; (b) twisted (Möbius) band; (c) torus; (d) Klein bottle.

(4.1) Theorem *Every compact, connected, orientable 2-manifold is homeomorphic to a sphere with handles added. Two such manifolds with the same number of handles are homeomorphic and conversely, so that the number of handles (called the genus) is the only topological invariant.*

This is a very satisfying theorem in that it shows that 2-manifolds of a certain large class can be enumerated and completely described to within homeomorphism (for a proof see Massey [1]). This can actually be carried further. Nonorientable as well as noncompact 2-manifolds can be described equally completely—although the noncompact case is more involved as might be expected. One can show also that every connected, one-dimensional manifold is homeomorphic to S^1 or to $\boldsymbol{R}$ depending on whether it is compact or not. However, beginning with $n = 3$ everything is far more complicated. In fact A. A. Markov has shown that for $n = 4$ no algorithm classifying compact orientable manifolds is possible (see Massey [1, p. 144]).

Curves and surfaces, that is, one- and two-dimensional manifolds in $\boldsymbol{E}^3$, formed the objects of study in classical differential geometry. We shall frequently refer to them as sources of examples and new ideas.

Exercises

1. Assuming invariance of domain, show that ∂H^n is a manifold of dimension $n - 1$ and that no neighborhood in H^n of a point of ∂H^n can be homeomorphic to an open subset of $\boldsymbol{R}^n$.
2. Prove that adding a handle to a 2-manifold in the fashion described above for S^2 and T^2 actually does give a 2-manifold.
3. Prove in detail that it is possible to obtain a 2-manifold by identifying sides of the square as shown in Fig. I.5d (Klein bottle).
4. Prove that identification of points at opposite ends of diameters on the boundary of the circular disk D^2 defines a 2-manifold.

According to a theorem of topology, if a compact orientable 2-manifold is obtained by pasting together triangles along their edges, then the number $\chi = f - e + v$ (faces $-$ edges $+$ vertices) is the same for two surfaces M_1 and M_2 which are homeomorphic: χ is independent of the way the surface is cut up into triangles. (χ is called the *Euler characteristic* of the surface.)

5. Let $M_0 = S^2$ and M_g be the surface obtained from M_0 by adding g handles. Compute the relation between the genus g and the Euler characteristic χ.

5 Abstract Manifolds. Some Examples

The manifolds of dimensions 1 and 2 considered above are pictured as *subspaces* of $\boldsymbol{E}^3$ except in the case of the Klein bottle. This is the way in which manifolds are first and most easily visualized. However, the definition

makes no such requirement. Such visualization makes equivalent (homeomorphic) manifolds *look* different just because they are differently placed in Euclidean space; and we might easily be led to think that they *are* different. Several examples of equivalent manifolds are shown in Fig. I.6. In spite of appearances, they are homeomorphic.

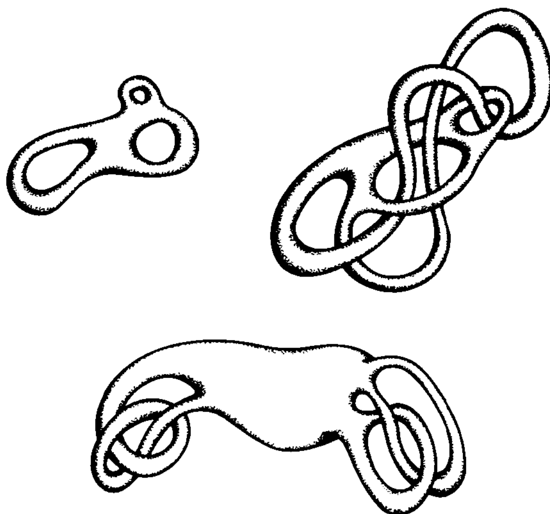


Figure I.6
Three equivalent manifolds.

As we might expect from the definition, it is possible to give examples of manifolds which we do *not* think of as lying in Euclidean space. Indeed, it is not clear that they can be realized at all as a subspace of Euclidean space. This can already be guessed from the construction of manifolds by pasting, which does not really use $\boldsymbol{E}^3$ at all. The simplest, as well as one of the most important examples of manifolds defined "abstractly"—not as a subspace of Euclidean space—is *real projective space* $P^n(\boldsymbol{R})$, the space of (real) projective geometry. It may be defined as follows. Let an equivalence relation $\sim$ be defined on $\boldsymbol{R}^{n+1} - \{0\}$ by

$$(x^1, \ldots, x^{n+1}) \sim (y^1, \ldots, y^{n+1})$$

if there is a real number t such that $y^i = tx^i$, $i = 1, \ldots, n + 1$; briefly $y = tx$. Then we denote by $[x]$ the equivalence class of x and by $P^n(\boldsymbol{R})$ the set of equivalence classes. There is a natural map $\pi : R^{n+1} - \{0\} \to P^n(\boldsymbol{R})$ given by $\pi(x) = [x]$ and we shall topologize $P^n(\boldsymbol{R})$, as is usual in the case of such quotient spaces, by saying that $U \subset P^n(\boldsymbol{R})$ is open if and only if $\pi^{-1}(U)$ is open in $\boldsymbol{R}^{n+1}$. This gives $P^n(\boldsymbol{R})$ the structure of an n-manifold (as shown in the exercises). We note that there is an alternative description of $P^n(\boldsymbol{R})$ as the

space of all lines through the origin 0 of $\boldsymbol{R}^{n+1}$; π takes each $x \neq 0$ to the line through 0 which contains it. Then we define the topology as follows: a collection $\tilde{U}$ of lines is open if it is the set of all lines through 0 which meet a given open set U.

This example may be generalized as follows: Let M be the set of all r-planes through the origin in $\boldsymbol{R}^n$, where n and r are fixed; for example, the set of all planes through the origin in $\boldsymbol{R}^3$ or the set of all three-dimensional planes through the origin of $\boldsymbol{R}^5$, and so on. This set has a natural topology which makes it a manifold. Intuitively it consists of defining a neighborhood of a given plane p to be all planes q which are "close" to it in a more or less obvious sense: there exist corresponding bases of both planes p and q (considered as r-dimensional subspaces of $\boldsymbol{R}^n$, as a vector space) such that corresponding basis vectors are close, say, for example, that their differences have norm less than some $\varepsilon > 0$.

Further important and useful examples of manifolds force themselves upon our attention when we begin to study the geometry of some of the manifolds we already have discussed. For example, consider S^2, the unit sphere in $\boldsymbol{R}^3$. We denote by $T(S^2)$ the collection of *all* tangent vectors to points of S^2, including the zero vector at each point. Thus $T(S^2) = \bigcup_{p \in S^2} T_p(S^2)$. This set has a natural topology: two tangent vectors X_p and Y_q are "close" if their initial points p and q *and* their terminal points are close. Similarly, if M is any of the 2-manifolds we have considered which lie "smoothly" in $\boldsymbol{E}^3$, so as to have a tangent plane at each point which turns continuously as we move about on M, then $T(M) = \bigcup_{p \in M} T_p(M)$ is a manifold, called the *tangent bundle of* M. The dimension of $T(M)$ is 4 since, roughly speaking, X_p depends locally on four parameters: two being the

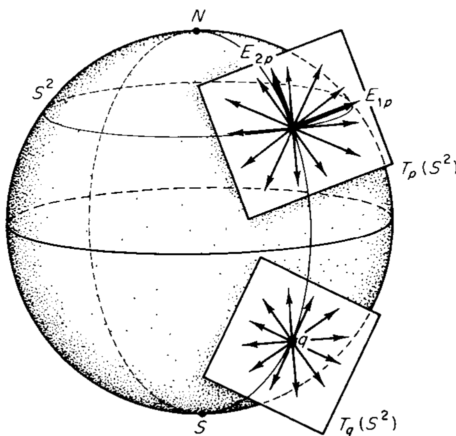


Figure I.7

The 2-sphere S^2 and some of its tangent vectors—elements of $T(S^2)$.

local coordinates of p relative to some coordinate neighborhood U and two more being the components which determine X_p relative to some basis $\{E_{1p}, E_{2p}\}$ of $T_p(M)$, a basis which varies continuously over the neighborhood U. We later make these statements quite precise and in so doing exhibit the locally Euclidean character of $T(M)$. For the moment we note that E_1 and E_2 can be visualized as vectors tangent to the coordinate curves $x^1 = \text{constant}$ and $x^2 = \text{constant}$ in U. This is illustrated in Fig. I.7.

We should note that these manifolds are *not* subspaces of $\boldsymbol{E}^3$, even though M is and although the geometry of $\boldsymbol{E}^3$ is used here to describe them. In fact, one of our major tasks is to describe $T_p(M)$ and $T(M)$ independently of any way of placing M in Euclidean space, that is, to give a description valid for an abstract manifold.

The manifolds mentioned above arose quite naturally from studies of the geometry of curves and surfaces in $\boldsymbol{E}^3$. In fact, Gauss used, in a very essential

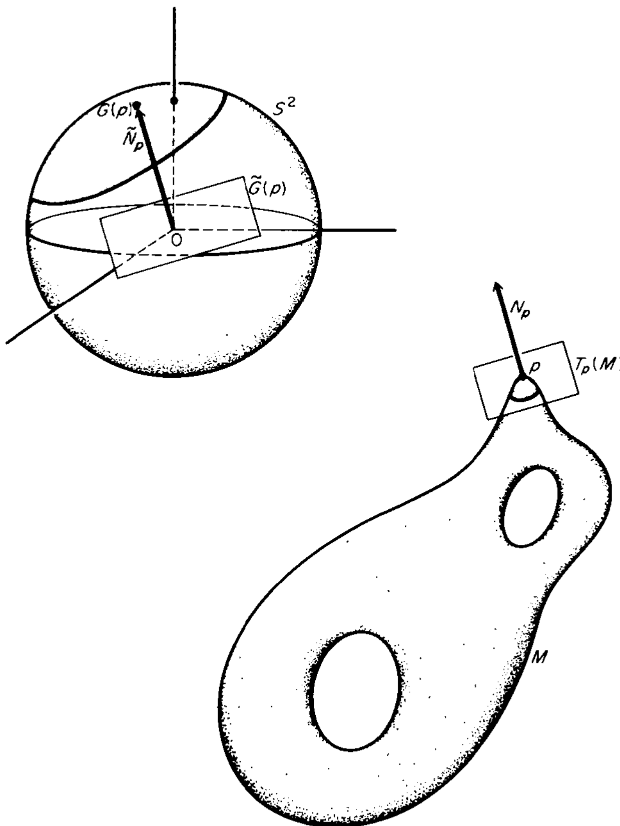


Figure I.8

way, a mapping which he introduced for orientable surfaces in $\boldsymbol{E}^3$. Let M be such a surface and let N_p be a unit normal vector at each $p \in M$, so defined that N_p varies continuously with p on M. Translate N_p to $\tilde{N}_p$ from a fixed origin 0 and let $G(p)$ be the endpoint of $\tilde{N}_p$ on S^2, the unit sphere at 0. The mapping taking p to $G(p)$ is known as the *Gauss mapping*, and the Gaussian curvature is a measure of the distortion of areas under this mapping: If M is sharply curved near p, then the area of a small region around p would be greatly magnified in mapping to S^2. Even if M is not orientable, we still have a tangent plane $T_p(M)$ at each point p and it is parallel to a uniquely determined plane $\tilde{G}(p)$ through the point 0. Thus a slight variant of the previous definition defines a mapping (as shown in Fig. I.8) of M to the manifold of 2-planes through 0 introduced above. Or again, using normal lines instead of tangent planes, we can obtain a mapping from M to the manifold of lines through 0, which as we have remarked, is equivalent to $P^2(\boldsymbol{R})$.

Exercises

1. Show that $P^2(\boldsymbol{R})$ and the manifold of Exercise 4.4 are homeomorphic.
2. Show that $P^2(\boldsymbol{R})$ and the set of all planes through the origin of $\boldsymbol{R}^3$ are in natural one-to-one correspondence.
3. Show that the set of all pairs $(\mathbf{x}, \mathbf{y})$ of mutually orthogonal unit vectors $\mathbf{x}$ and $\mathbf{y}$ of V^3, with its natural inner product, is a manifold. What is its dimension? Generalize if possible.
4. Prove that the manifold of orthonormal pairs of vectors in V^3 (Exercise 3) is homeomorphic to $T_0(S^2)$, the tangent sphere bundle of S^2, which consists of all unit vectors tangent to S^2.
5. Let C be a one-dimensional manifold (curve) in $\boldsymbol{R}^3$. Show that the collection of all vectors normal to C form a three-dimensional manifold. What sort of manifold would the unit vectors normal to C give us?
6. Manifolds may be obtained as the locus of one or more algebraic equations, for example, $S^2 = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$. Show that the torus T^2 may be given as the locus of an equation in x, y, z.

Notes

Curves and surfaces in Euclidean space were studied since the earliest days of geometry and, after they were invented, both analytic geometry and calculus were systematically used in these studies. However, the discoveries of Gauss, announced in 1827, profoundly altered the course of differential geometry and pointed the way to the concept of abstract differentiable manifolds—the underlying spaces of every geometry and of other important mathematical theories as well. In his celebrated "Theorema Egregium" Gauss showed that there is a measure of curvature of a surface (now called the *Gaussian curvature*) which depends only on one's ability to measure the lengths of curves on the surface. This means that this curvature is unchanged by alterations of shape of the surface which leave arclength unchanged. (It is easily

seen that there are such alterations. For example, we may roll a plane surface into a cylinder or cone, or we may gently squeeze a hemisphere in along its edge, the equator.) This discovery of an "inner" geometry, independent of the shape of the surface in E^3, led very naturally toward the invention of abstract surfaces (2-manifolds) on which a measure of arclength is (somehow) provided. The discovery by Bolyai and Lobachevskii (independently) about 1830 of non-Euclidean geometry was another important step toward the concept of abstract spaces—spaces not conceived of as subspaces of Euclidean space. (Non-Euclidean geometry satisfies all of Euclid's postulates except the one which affirms that through any point p not on a line L there is exactly one line parallel to L. As in Euclidean geometry, lengths of curves and distances between points have meaning; see hyperbolic space, Section VIII.6.) The existence of such spaces was apparently already known to Gauss, who kept his knowledge to himself—presumably for fear of the profoundly disturbing philosophical and religious consequences such a discovery would bring about. It can only be compared with the discovery of relativity in this century, for which it laid the foundation. The belief that we live in a Euclidean world had been an article of faith for centuries and was not easily challenged.

A second great impetus to these new ideas was given by Riemann in his inaugural address at Göttingen in 1854. He introduced the idea of a space or "manifold" having its existence outside of Euclidean space but locally like it; he made clear what arc length would mean in this case (see Section V.3)—separating metric and topological properties. Both Euclidean and non-Euclidean geometry and spaces of arbitrary dimension were included within his ideas. Later he made extensive use of the notion of abstract two-dimensional manifolds in analytic function theory by his systematic use of Riemann surfaces.

In the remainder of the 19th century many further important developments in geometry took place. The most important by far from the point of view of modern differential geometry and manifold theory being the monumental work of Poincaré. He introduced concepts and raised questions which have dominated much of the work in these fields in this century. Among them the use of groups to measure topological invariants of manifolds, the importance of manifolds in the study of dynamical systems, the Poincaré conjecture—central to an understanding of three dimensional manifolds but still unresolved, and many others. A second great impetus coming at the end of the 19th century and early in this century was the work on Lie groups, especially in the hands of the great differential geometer E. Cartan. In his famous Erlangen program of 1872 Felix Klein had shown the intimate relation of all geometries with groups, and with the later development of Lie groups (themselves important differentiable manifolds) it became clear that both Euclidean and non-Euclidean geometry could be thought of as branches of group theory. All of these applications gradually clarified the concepts themselves, as did the emergence of topology, so that the ideas of manifold theory and differential geometry are now highly developed and used across the entire mathematical spectrum, in relativity theory, analysis, Lie groups, algebraic topology, algebraic geometry, and elsewhere. The reader will find historical sketches in many of the references. In particular, Gauss's famous paper [1] is available in an annotated English translation and Riemann's Inaugural Address is translated in the book of Spivak [2], which contains an excellent and very readable analysis of this paper and is highly recommended for those interested in a history of differential geometry. The reader will also find an elegant intuitive discussion of surfaces given by Hilbert and Cohn-Vossen [1].

II FUNCTIONS OF SEVERAL VARIABLES AND MAPPINGS

In this chapter we review in some detail the differential calculus which we will need later. The purpose is to build a bridge between the reader's previous knowledge of multivariable calculus and the somewhat specialized facts we need here, especially the inverse function theorem and the theorem on rank. (Many readers can skim over or skip this chapter entirely.)

Briefly, the topics treated are the following: In Section 1 we define differentiability of real-valued functions of many variables and its immediate consequences, in particular the mean value theorem. In Section 2 this is extended to the case that concerns us most, a mapping F from an open subset U of $\boldsymbol{R}^n$ into $\boldsymbol{R}^m$. Here the Jacobian is defined and the mean value theorem restated for mappings. Sections 3 and 4 deal with the concept of the space of tangent vectors $T_a(\boldsymbol{R}^n)$ at a point $a \in \boldsymbol{R}^n$; this will be most important in studying manifolds, especially Section 4 in which $T_a(\boldsymbol{R}^n)$ is defined in a way that admits generalization. Section 5 reviews the notion of vector field in $\boldsymbol{R}^n$. Section 6 gives a detailed proof of the inverse function theorem. This theorem with its corollaries, especially the theorem on rank (Section 7), is one of the basic theorems on which most of our theory is built.

1 Differentiability for Functions of Several Variables

In this section we review briefly some facts about partial derivatives from advanced calculus. Few proofs are given; they may be worked out as problems or found in advanced calculus texts, for example, Apostol [1],

Dieudonné [1], or Fleming [1]. We will consider real-valued functions of several variables, more precisely functions whose domain is a subset $A \subset \boldsymbol{R}^n$ and whose range is $\boldsymbol{R}$. If $f\colon A \to \boldsymbol{R}$ is such a function, then $f(x) = f(x^1, \ldots, x^n)$ denotes its value at $x = (x^1, \ldots, x^n) \in A$. We assume throughout this section that f is a function on an *open* set $U \subset \boldsymbol{R}^n$. At each $a \in U$, the *partial derivative* $(\partial f/\partial x^j)_a$ of f with respect to x^j is, of course, the following limit, if it exists:

$$\left(\frac{\partial f}{\partial x^j}\right)_a = \lim_{h \to 0} \frac{f(a^1, \ldots, a^j + h, \ldots, a^n) - f(a^1, \ldots, a^j, \ldots, a^n)}{h}.$$

If $\partial f/\partial x^j$ is defined, that is, the limit above exists at each point of U for $1 \le j \le n$, this defines n functions on U. Should these functions be continuous on U for $1 \le j \le n$, f is said to be *continuously differentiable* on U, denoted by $f \in C^1(U)$.

Mere existence of partial derivatives is too weak a property for most purposes. For example, the function defined on $\boldsymbol{R}^2$ by

$$f(x, y) = \frac{xy}{x^2 + y^2} \qquad \text{and} \qquad f(0, 0) = 0$$

is not continuous at $(0, 0)$, yet both derivatives are defined there. The natural generalization of existence of the derivative for functions of one variable is as follows. We shall say that f is *differentiable* at $a \in U$ if there is a (homogeneous) linear expression $\sum_{i=1}^n b_i(x^i - a^i)$ such that the (inhomogeneous) linear function defined by $f(a) + \sum_{i=1}^n b_i(x^i - a^i)$ approximates $f(x)$ near a in the following sense:

$$\lim_{x \to a} \frac{f(x) - f(a) - \sum b_i(x^i - a^i)}{\|x - a\|} = 0,$$

or equivalently, if there exist constants $b_1, \ldots, b_n$ and a function $r(x, a)$ defined on a neighborhood V of $a \in U$ which satisfy the following two conditions:

$$f(x) = f(a) + \sum b_i(x^i - a^i) + \|x - a\| r(x, a)$$

on V, and

$$\lim_{x \to a} r(x, a) = 0.$$

If f is differentiable for every $a \in U$, we say it is *differentiable on U*. [*Warning:* this is a technical definition from advanced calculus. Beginning with Chapter III *differentiable* will be used rather loosely to mean differentiable of some order, usually infinitely differentiable (C^∞).] Note that differentiability on U is a local concept, that is, if f is differentiable on a neighborhood of each point of U, then f is differentiable on U. By the mean value theorem, for

a function of one variable the existence of the derivative at $a \in U$ is equivalent to differentiability; but for functions of several variables, as we have seen, this is not the case. The exercises at the end of this section and the following statements (1.1)–(1.3), whose proofs we leave as exercises, will clarify these concepts.

(1.1) *If f is differentiable at a, then it is continuous at a and all the partial derivatives $(\partial f/\partial x^i)_a$ exist. Moreover the b_i are uniquely determined for each a at which f is differentiable; in fact $b_i = (\partial f/\partial x^i)_a$.*

By virtue of (1.1) when f is differentiable at a we have

$$f(x) - f(a) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i}\right)_a (x^i - a^i) + \|x - a\| r(x, a).$$

We denote by $(df)_a$, or simply df, the homogeneous linear expression on the right:

$$df = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i}\right)_a (x^i - a^i). \tag{1.2}$$

It is called the differential of f at a.

(1.3) *If $\partial f/\partial x^1, \ldots, \partial f/\partial x^n$ are defined in a neighborhood of a and continuous at a, then f is differentiable at a.*

Thus existence and continuity of the partial derivatives of f on an open set $U \subset \boldsymbol{R}^n$ implies differentiability of f at every point of U. We define inductively the notion of an *r-fold continuously differentiable function* on an open set $U \subset \boldsymbol{R}^n$ (function of class C^r): f is of class C^r on U if its first derivatives are of class C^{r-1}. Equivalently we may say that f has *continuous derivatives of order* $1, 2, \ldots, r$ on U. If f is of class C^r for all r, then we say that f is *smooth*, or of class C^∞. As in the case of C^1, we denote these classes of functions on U by $C^r(U)$ and $C^\infty(U)$.

We now state the first version of the chain rule; a more general version will be given in the next section. Define a *differentiable* (C^r) *curve* in $\boldsymbol{R}^n$ to be a mapping of an open interval $(a, b) = \{x \in \boldsymbol{R} \mid a < x < b\}$ of the real numbers into $\boldsymbol{R}^n$, $f\colon (a, b) \to \boldsymbol{R}^n$, with $f(t) = \{x^1(t), \ldots, x^n(t)\}$, where the n coordinate functions $x^1(t), \ldots, x^n(t)$ are differentiable (resp. C^r) on the interval. (Recall: For functions of *one* variable "differentiable" and "derivative exists" are equivalent.) Now suppose that f is a differentiable curve and maps (a, b) into U, an open subset of $\boldsymbol{R}^n$. Let $a < t_0 < b$ and suppose that g is a function on U which is differentiable at $f(t_0) \in U$. Then the composite function $g \circ f$ is a

real-valued function on (a, b). We assert that $g \circ f$ is differentiable at t_0 and that its derivative at t_0 is given by the *chain rule*

$$\frac{d}{dt}(g \circ f)_{t_0} = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x^i}\right)_{f(t_0)} \left(\frac{dx^i}{dt}\right)_{t_0}. \tag{1.4}$$

The proof is left as an exercise. Using it we may establish the mean value theorem for functions of several variables. We shall say that a domain U is *starlike* with respect to $a \in U$ provided that whenever $x \in U$, then the segment $\overline{ax}$ lies entirely in U (see Fig. II.1). This is a somewhat weaker property than convexity of U, a convex set being starlike with respect to every one of its points.

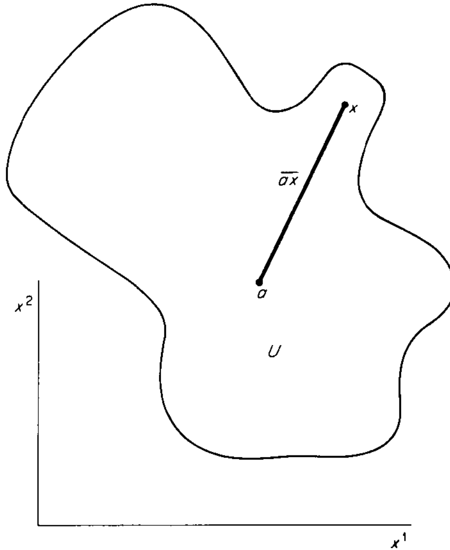


Figure II.1

(1.5) Theorem (Mean Value Theorem) *Let g be a differentiable function on an open set $U \subset \boldsymbol{R}^n$; let $a \in U$ and suppose that U is starlike with respect to a. Then given $x \in U$ there exists $\theta \in \boldsymbol{R}$, $0 < \theta < 1$, such that*

$$g(x) - g(a) = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x^i}\right)(x^i - a^i),$$

the derivatives $\partial g/\partial x^1, \ldots, \partial g/\partial x^n$ all being evaluated at the same point $a + \theta(x - a)$ on the segment $\overline{ax}$.

Proof Set $f(t) = a + t(x - a)$, that is $x^i(t) = a^i + t(x^i - a^i)$. Then the corresponding curve is a line segment with $f(0) = a$ and $f(1) = x$. This curve

is differentiable, in fact C^∞, so that $g \circ f$ maps $[0, 1]$ into U and is differentiable on $(0, 1)$. Applying the standard mean value theorem for functions of one variable (as in elementary differential calculus) and using (1.4) to compute the derivatives gives the formula. ∎

(1.6) Corollary *Let U and g be as in Theorem 1.5. If $|\partial g/\partial x^i| < K$ on U, $i = 1, 2, \ldots, n$, then for any $x \in U$, we have*

$$|g(x) - g(a)| < K\sqrt{n}\,\|x - a\|.$$

Proof Taking absolute values in the formula of Theorem 1.5 and using the Schwarz inequality gives

$$|g(x) - g(a)| = \left|\sum_{i=1}^{n}\left(\frac{\partial g}{\partial x^i}\right)(x^i - a^i)\right| \leq \left[\sum_i \left(\frac{\partial g}{\partial x^i}\right)^2\right]^{1/2}\left[\sum_i (x^i - a^i)^2\right]^{1/2}.$$

Therefore

$$|g(x) - g(a)| < K\sqrt{n}\,\|x - a\|.$$ ∎

The following corollary is an important consequence and should be proved as an exercise.

(1.7) Corollary *If f is of class C^r on U, then at any point of U the value of the derivatives of order k, $1 < k \leq r$ is independent of the order of differentiation, that is, if $(j_1, \ldots, j_k)$ is a permutation of $(i_1, \ldots, i_k)$, then*

$$\frac{\partial^k f}{\partial x^{i_1} \cdots \partial x^{i_k}} = \frac{\partial^k f}{\partial x^{j_1} \cdots \partial x^{j_k}}.$$

Taylor's theorem on polynomial approximation with its formula for the remainder R_{N+1}, or error of the approximation of degree N, as well as the corollary theorem on power series expansions are easily extended to functions of several variables using the technique of Theorem 1.5 (see Apostol [1] and Dieudonné [1]). As in the single variable case, a necessary but not a sufficient condition that a function be (real) analytic, that is, can be expanded in a power series at each $a \in U$, an open set of $\boldsymbol{R}^n$, is that it be in $C^\infty(U)$. [We write $f \in C^\omega(U)$ if f is real analytic on U.] Although knowledge of analytic functions is not needed in this text, it is helpful—since C^ω implies C^∞—to know that any linear function $f(x) = \sum a_i x^i$, or any polynomial $P(x^1, \ldots, x^n)$ of n variables, is an analytic function on $U = \boldsymbol{R}^n$; the same is true for any quotient of polynomials (rational function) if we exclude from the domain the points at which the denominator is zero. Thus, for example, a determinant is an analytic function of its entries and, if we exclude $n \times n$ matrices of determinant zero (which have no inverses), then each entry in the inverse A^{-1} of a matrix a is an analytic (and hence C^∞) function of the entries in the matrix A.

Exercises

1. Prove (1.1).
2. Prove (1.3) using the mean value theorem for functions of one variable.
3. Prove that all first partial derivatives of a differentiable function vanish at an extremum.
4. Let U be an open subset of $\boldsymbol{R}^n$ and let $C^0(U)$ and $C^1(U)$ denote the continuous and continuously differentiable functions on U. Let $D(U)$ denote the functions which are differentiable on U. Show that $C^0(U) \supset D(U) \supset C^1(U)$ and construct examples to show that in general the inclusions are proper.
5. Show that the inclusions $C^1(U) \supset C^2(U) \supset \cdots \supset C^\infty(U)$ are proper.
6. Prove that $C^\infty \supset C^\omega$, and that the inclusion is proper. [*Hint:* Let $f(0) = 0, f(t) = \exp(-1/t^2)$ for $t \neq 0$; f is C^∞ on $\boldsymbol{R}$. Is it C^ω on $\boldsymbol{R}$?]
7. Prove (1.4), that is, prove that $g \circ f$ is differentiable at $t = t_0$ and that its derivative is given by (1.4).
8. Prove Corollary 1.7.

Sometimes it is important to extend the definitions of differentiability, C^1, and so on, to functions defined on a subset $A \subset \boldsymbol{R}^n$, which is not assumed to be an open set, for example, a function $f(t)$ of one variable on the closed interval $0 \leq t \leq 1$. We say that f is *differentiable*, of class C^r, of class C^∞, and so on, on A if f can be extended to a differentiable, C^r, C^∞ function, respectively, on an open subset U of $\boldsymbol{R}^n$ which contains A.

9. Show that if $A = \{x \in \boldsymbol{R}^n \mid a^i \leq x^i \leq b^i, a^i < b^i, i = 1, \ldots, n\}$ and f is differentiable on A, then the value of $\partial f/\partial x^i$ at any point of A is independent of the extension chosen. Can you find any example to show that for some sets A this is not the case? If so, does assuming C^1 help?

2 Differentiability of Mappings and Jacobians

In this section we generalize the ideas of the previous section to the case of functions defined on subsets of $\boldsymbol{R}^n$ but whose range is in $\boldsymbol{R}^m$ rather than $\boldsymbol{R}$. We will refer to them as *mappings* (or maps) and, insofar as possible, reserve the term *function* for real-valued functions as in Section 1. If $\pi^i\colon \boldsymbol{R}^m \to \boldsymbol{R}$ denotes the projection to the ith coordinate, namely, $\pi^i(x^1, \ldots, x^i, \ldots, x^m) = x^i$, and if $F\colon A \to \boldsymbol{R}^m$ is a mapping defined on $A \subset \boldsymbol{R}^n$, then F is determined by its *coordinate functions* $f^i = \pi^i \circ F$; in fact for $x \in A$,

$$F(x) = (f^1(x), \ldots, f^m(x)).$$

Conversely, any set of m functions $f^1, \ldots, f^m$ on A with values in R determines a mapping $F\colon A \to \boldsymbol{R}^m$ with the coordinates of $F(x)$ given by $f^1(x), \ldots, f^m(x)$ as above.

We are interested in the case where U is an open set of $\boldsymbol{R}^n$, possibly all of $\boldsymbol{R}^n$. Since many authors identify $\boldsymbol{R}^m$ and V^m (see Section I.1), these are sometimes referred to as *vector-valued functions on* $\boldsymbol{R}^n$, although we will not use that terminology. From general topology we know that F is continuous if and only if its coordinate functions are. We shall say that F is *differentiable, of class* C^r, C^∞, C^ω, and so on, *at* $a \in U$ or *on* U if each of its coordinate functions has the corresponding property. We may sometimes call a C^∞ mapping F a *smooth* mapping; if F is smooth, then each coordinate function f^i possesses continuous partial derivatives of all orders and each such derivative is independent of the order of differentiation.

If F is differentiable on U, we know that the $m \times n$ *Jacobian matrix*

$$\frac{\partial(f^1, \ldots, f^m)}{\partial(x^1, \ldots, x^n)} = \begin{pmatrix} \dfrac{\partial f^1}{\partial x^1} & \cdots & \dfrac{\partial f^1}{\partial x^n} \\ \vdots & & \vdots \\ \dfrac{\partial f^m}{\partial x^1} & \cdots & \dfrac{\partial f^m}{\partial x^n} \end{pmatrix}$$

is defined at each point of U, its mn entries being functions on U. These functions need not be continuous on U; they are so if and only if F is of class C^1. Since differentiability is needed in Section II.6, it is useful to give another formulation of this concept for mappings. We leave the proof to the exercises. [*Note:* later *differentiable* will be used to mean C^∞; see Section III.3.]

(2.1) *A mapping* $F: U \to \boldsymbol{R}^m$, *$U$ an open subset of* $\boldsymbol{R}^n$, *is differentiable at* $a \in U$ *(or on U) if and only if there exists an $m \times n$ matrix A of constants (respectively, functions on U) and an m-tuple* $R(x, a) = (r^1(x, a), \ldots, r^m(x, a))$ *of functions defined on U (on $U \times U$) such that* $\|R(x, a)\| \to 0$ as $x \to a$ *and for each $x \in U$ we have*

$$(*) \qquad F(x) = F(a) + A(x - a) + \|x - a\| R(x, a).$$

If such $R(x, a)$ and A exist, then A is unique and is the Jacobian matrix.

[In the expression $(*)$, $A(x - a)$ denotes a matrix product, so we must write $(x - a)$ as an $n \times 1$ (column) matrix and read this as an equation in $m \times 1$ matrices. The last term means that each component of the m-tuple $R(x, a)$ is multiplied by $\|x - a\|$.]

Corollary 1.6 extends immediately to mappings in the following form. The proof is left as an exercise.

(2.2) Theorem *Let $a \in U$ be an open subset of* $\boldsymbol{R}^n$ *which is starlike with respect to a, and let* $F: U \to \boldsymbol{R}^m$ *be differentiable on U with* $|\partial f^i/\partial x^j| \leq K$,

$1 \le i \le m$, $1 \le j \le n$, at every point of U. Then the following inequality holds for all $x \in U$:

$$(**) \qquad \|F(x) - F(a)\| \le (nm)^{1/2} K \|x - a\|.$$

We will use DF to denote the Jacobian matrix of a differentiable mapping F and $DF(x)$ to denote its value at x. If F is differentiable on U, then for $a \in U$ expression $(*)$ becomes

$$F(x) = F(a) + DF(a)(x - a) + \|x - a\| R(x, a).$$

We remark that $F \in C^1(U)$ if and only if $DF(x)$ varies continuously with x, that is, $x \to DF(x)$ is a continuous map of U into the space $\mathscr{M}_{mn}(\boldsymbol{R})$ of $m \times n$ matrices, identified with $\boldsymbol{R}^{mn}$ and given the corresponding topology.

Just as in the case of functions we wish to prove a chain rule for composition of mappings. We suppose $U \subset \boldsymbol{R}^n$ and $V \subset \boldsymbol{R}^m$ are open sets and $F\colon U \to V \subset \boldsymbol{R}^m$ and $G\colon V \to \boldsymbol{R}^p$ so that $H = G \circ F$ is defined on U, which it maps into $\boldsymbol{R}^p$. We may write the coordinate functions of H using those of F and G:

$$h^i(x) = g^i \circ F(x) = g^i(f^1(x), \ldots, f^m(x)), \qquad i = 1, \ldots, p.$$

(2.3) Theorem (Chain rule) *Let F, G, H be as above and suppose that F is differentiable at $a \in U$ and G is differentiable at $b = F(a)$. Then $H = G \circ F$ is differentiable at $x = a$ and we have*

$$DH(a) = DG(F(a)) \cdot DF(a)$$

(where $\cdot$ indicates matrix multiplication). If F is differentiable on U and G on V, then this holds for every $a \in U$.

Proof According to the characterization above it is enough to show that the p-tuple $R_H(x, a)$ defined by

$$H(x) - H(a) - DG(F(a)) \cdot DF(a) \cdot (x - a) = \|x - a\| R_H(x, a)$$

approaches 0 as x approaches a. Using $y = F(x)$, $b = F(a)$, and the differentiability of F and G at a and b, we may write

$$H(x) - H(a) = G(y) - G(b) = DG(b) \cdot (y - b) + \|y - b\| R_G(y, b),$$

and

$$y - b = F(x) - F(a) = DF(a) \cdot (x - a) + \|x - a\| R_F(x, a).$$

Then, replacing y by $F(x)$ and b by $F(a)$,

$$H(x) - H(a) = DG(a) \cdot DF(a) \cdot (x - a)$$
$$+ \|x - a\| \left\{ DG(F(a)) \cdot R_F(x, a) + \frac{\|F(x) - F(a)\|}{\|x - a\|} R_G(F(x), F(a)) \right\}.$$

Using the continuity of F, which is an immediate consequence of differentiability, and the properties of $R_F(x, a)$ and $R_G(y, b)$, we see that as $x \to a$ the expression in curly braces, which we may denote by $R_H(x, a)$, goes to zero. This completes the proof. ∎

(2.4) Corollary *If F and G are of class C^r (or smooth) on U and V, respectively, then $H = G \circ F$ is of class C^r (or smooth) on U.*

Proof We prove only the statement for C^1, although we will use the general case, whose proof is a problem in mathematical induction (see Dieudonné [1], where it is also proved for analytic mappings). If F and G are C^1, then they are certainly differentiable, and DF and DG are continuous functions on U and V. Since F is C^1, it is continuous and so $DG(F(x))$ is continuous on U. Finally the product of two matrices is a continuous, in fact C^ω, mapping of $\boldsymbol{R}^{mn} \times \boldsymbol{R}^{np}$ since the entries in the $m \times p$ product matrix are polynomials in the entries of the factors. Thus the chain rule formula gives $DH(x)$ as a composite of functions which are at least continuous so that it must be continuous. This is equivalent to its entries being continuous which means that the coordinate functions of H, and thus H itself, are of class C^1. ∎

Exercises

1. Prove (2.1).
2. Prove Theorem 2.2.
3. Prove Corollary 2.4 for the case $r = 2$ and try to construct a procedure which would give the result for all r.
4. Prove that the rank of the product of two matrices is less than or equal to the rank of either factor. Show that multiplying a matrix on the left or right by a nonsingular matrix does not change its rank.

Just as in the case of functions, we can extend the notion of differentiability, C^r, C^∞, and so on, to mappings into $\boldsymbol{R}^m$ whose domain of definition is an *arbitrary* subset $A \subset \boldsymbol{R}^n$. We say that $F: A \to \boldsymbol{R}^m$ is *differentiable*, C^r, C^∞, or C^ω if and only if it has an extension to an open set $U \supset A$ which is, respectively, differentiable, C^r, C^∞, or C^ω. As we have mentioned in Section 1, there exist examples to show that under these circumstances $DF(x)$ may not be uniquely defined at each $x \in A$, that is, it may depend on the extension of F, the simplest example being that A is a single point. Thus one must use some care in dealing with this case. The following two problems involve this generalization.

5. Let U be an open subset of $\boldsymbol{R}^n$ and $F: U \to \boldsymbol{R}^m$, $m \leq n$, be a C^1 mapping. Suppose that F is injective (one-to-one into) and that $F^{-1}: A \to U$, where $A = F(U)$ is also of class C^1. Then show that m cannot be less than n. (This is a weak version of the theorem of Brouwer: There exists no homeomorphism of an open set U of $\boldsymbol{R}^n$ into $\boldsymbol{R}^m$, $m < n$.)

6. Let A be a closed cube in $\boldsymbol{R}^3$ and suppose $F: A \to \boldsymbol{R}^m$ is a mapping of class C^1. Prove that the value of DF on A is independent of any extension. Generalize this to other domains A and to class C^∞.

3 The Space of Tangent Vectors at a Point of R^n

Although we shall presently restrict our attention to $\boldsymbol{R}^n$, let us first consider $\boldsymbol{E}^n$, or $\boldsymbol{E}^3$ at least, for the sake of intuition. Our purpose is to attach to *each* point a of $\boldsymbol{R}^n$ an n-dimensional vector space $T_a(\boldsymbol{R}^n)$. We know how to do this in Euclidean space: If $a \in \boldsymbol{E}^3$, we let $T_a(\boldsymbol{E}^3)$ be the vector space whose elements are directed line segments X_a with a as initial point. These are added by the parallelogram law: $-X_a$ is the oppositely directed segment and 0 is the segment consisting of the point a alone. We have supposed that a unit of length was chosen in $\boldsymbol{E}^3$ and we may denote by $\|X_a\|$ the length of the segment. Multiplication by positive (negative) real numbers leaves the direction unchanged (reversed) and multiplies the length by the absolute value of the number. To show that this does indeed give a vector space of dimension 3 over $\boldsymbol{R}$ is an exercise in solid geometry. Thus we attach to *each* point of $\boldsymbol{E}^3$ a three-dimensional vector space called the *tangent space* at that point.

We shall ultimately attach vector spaces at each point of more complicated spaces, namely manifolds; this was briefly indicated in Section I.5. There is, however, a unique feature of the tangent spaces of Euclidean space which is not shared by the tangent spaces at points of manifolds; the tangent spaces at any two points of Euclidean space are *naturally isomorphic*, that is, there is an isomorphism determined in some unique fashion by the geometry of the space—not chosen by us. (Without the restriction of naturality, the statement would be trivial since any two vector spaces of the same dimension over the real numbers are isomorphic, but in general there is no unique isomorphism singled out, rather we must choose one arbitrarily from a very large collection.)

Indeed, if a, b are points of $\boldsymbol{E}^3$, then there is exactly one translation of the space taking a to b; this translation moves each line segment issuing from a to a line segment from b and thus carries $T_a(\boldsymbol{E}^3)$ to $T_b(\boldsymbol{E}^3)$. Since parallelograms go to congruent parallelograms and lengths are preserved, this correspondence is an isomorphism; and it is uniquely determined by the geometry (Fig.II.2). If we choose a fixed point a as origin and choose at a three linearly independent vectors E_{1a}, E_{2a}, E_{3a}, for example, three mutually perpendicular unit vectors, then this will automatically determine a basis not only of $T_a(\boldsymbol{E}^3)$ but also (by parallel translation) of $T_b(\boldsymbol{E}^3)$ for every $b \in \boldsymbol{E}^3$. All of this is intuitive geometry and we have not really proved the statements we have made. Therefore we turn to $\boldsymbol{R}^n$ where we are able to be more precise and rigorous, but we keep in mind our geometric model.

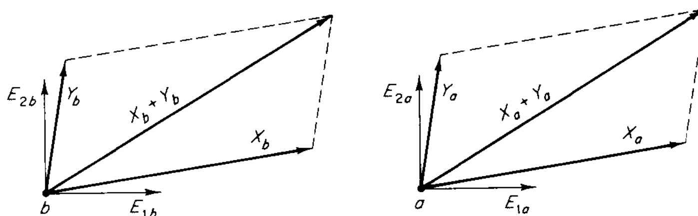


Figure II.2

Let $a = (a^1, \ldots, a^n)$ be any point of $\boldsymbol{R}^n$. We define $T_a(\boldsymbol{R}^n)$, the *tangent* (vector) *space* attached to a, as follows. First, as a *set* it consists of all pairs of points (a, x), or $\overrightarrow{ax}$, $a = (a^1, \ldots, a^n)$ and $x = (x^1, \ldots, x^n)$, corresponding, of course, to initial and terminal points of a segment. We also denote such a pair by X_a, using upper case letters for vectors. We next establish a one-to-one correspondence $\varphi_a \colon T_a(\boldsymbol{R}^n) \to V^n$ between the set just described and the vector space of n-tuples of real numbers by the following simple device: If $X_a = \overrightarrow{ax}$, then $\varphi_a(X_a) = (x^1 - a^1, \ldots, x^n - a^n)$. Finally the *vector space operations* (addition and multiplication by scalars) are defined in the one way possible so that φ_a is an isomorphism. This requires that

$$X_a + Y_a = \varphi_a^{-1}(\varphi_a(X_a) + \varphi_a(Y_a)),$$
$$\alpha X_a = \varphi_a^{-1}(\alpha\varphi_a(X_a)), \qquad \alpha \in \boldsymbol{R},$$

the right-hand side being used to define the operations on the left. Clearly we are being guided by the fact that $\boldsymbol{R}^n$ and $\boldsymbol{E}^n$ may be identified if we choose rectangular Cartesian coordinates in $\boldsymbol{E}^n$. This is equivalent to choosing an origin 0 and n mutually orthogonal unit vectors there, $(E_1)_0, \ldots, (E_n)_0$, lying on each (positive) coordinate axis—as do **i**, **j**, and **k** in the usual model for $\boldsymbol{E}^3$. Then vectors at any point a are uniquely determined by their components relative to the basis $E_{1a}, \ldots, E_{na}$, which in turn are given by subtracting from the coordinates of the terminal point of each vector, the coordinates of its initial point a. The geometry of $\boldsymbol{E}^n$ has guided us to a proper method for defining the tangent space at each point of $\boldsymbol{R}^n$. Please note that V^n has a canonical basis $\mathbf{e}^1 = (1, 0, \ldots, 0), \ldots, \mathbf{e}^n = (0, \ldots, 1)$ and this gives at each $a \in \boldsymbol{R}^n$ a *natural* or *canonical basis* $E_{1a} = \varphi_a^{-1}(\mathbf{e}_1), \ldots, E_{na} = \varphi_a^{-1}(\mathbf{e}_n)$ of $T_a(\boldsymbol{R}^n)$. The canonical isomorphism given by translation in the case of $\boldsymbol{E}^n$ is now $\varphi_b^{-1} \circ \varphi_a \colon T_a(\boldsymbol{R}^n) \to T_b(\boldsymbol{R}^n)$, and we have $X_a = \overrightarrow{ax}$ corresponding to $Y_b = \overrightarrow{by}$ if and only if $x^i - a^i = y^i - b^i$, $i = 1, 2, \ldots, n$. However, we *never* identify the tangent spaces into a single vector space as is often done in discussions of vectors on Euclidean space, that is, we never equate vectors

with *different* initial points; in particular, we cannot add a vector in $T_a(\boldsymbol{R}^n)$ and one in $T_b(\boldsymbol{R}^n)$ where $a \neq b$. The reason for our insistence on this point will appear when we learn how to attach a tangent space $T_p(M)$ to each point p of a manifold in general, for then we have nothing corresponding to the natural isomorphisms of $T_a(\boldsymbol{E}^3)$ and $T_b(\boldsymbol{E}^3)$ given by the translations of $\boldsymbol{E}^3$. For example, there is no natural isomorphism of the tangent vectors to S^2 at two distinct points p and q of S^2. It follows that our method of defining $T_a(\boldsymbol{R}^n)$ at each a—which depended on such an isomorphism—is not suitable for generalization in its present form. Therefore we shall give two further methods for defining $T_a(\boldsymbol{R}^n)$, one in this section with details left as exercises and a second, which we use in the remainder of the text, in the section following this one.

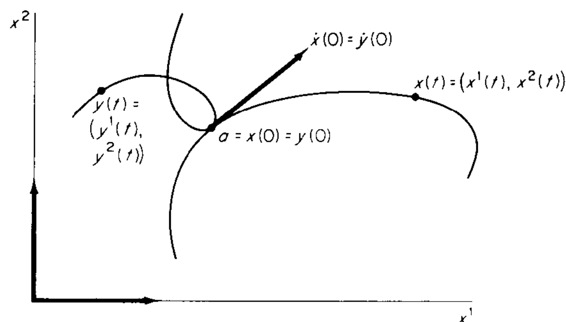


Figure II.3
Equivalent Curves: $x(t)$ and $y(t)$.

We begin by a formal description of the first definition. Let $x(t)$, $-\varepsilon < t < \varepsilon$, be a C^1 curve in $\boldsymbol{R}^n$ passing through $a \in \boldsymbol{R}^n$ when $t = 0$, that is, assume $x(t) = (x^1(t), \ldots, x^n(t))$, where $x^i(t)$ is C^1 and $x^i(0) = a^i$, $i = 1, \ldots, n$. Let $I_\varepsilon = \{t \in \boldsymbol{R} \mid |t| < \varepsilon\}$. Then each such curve is a C^1 map of $I_\varepsilon \to \boldsymbol{R}^n$, where $\varepsilon > 0$ and may vary from curve to curve. Two curves are *equivalent*, $x(t) \sim y(t)$, if at $t = 0$, the derivatives with respect to t of their coordinate functions are equal: $\dot{x}^i(0) = \dot{y}^i(0)$, $i = 1, \ldots, n$. Let $[x(t)]$ denote the equivalence class of $x(t)$; to each $[x(t)]$ corresponds an n-tuple of numbers $\dot{x}(0) = (\dot{x}^1(0), \ldots, \dot{x}^n(0))$, that is, an element of $\boldsymbol{V}^n$. Using this map we obtain a vector space structure on the collection of equivalence classes which we denote, predictably, by $T_a(\boldsymbol{R}^n)$. Details are left as exercises. Intuitively speaking, if we use the identification of $\boldsymbol{R}^n$ with $\boldsymbol{E}^n$ plus a rectangular Cartesian coordinate system, we see that $\dot{x}^i(0)$ is the ith component of the velocity vector of the particle whose motion is given by $x(t) = (x^1(t), \ldots, x^n(t))$ at the instant it passes through a (see Fig. II.3). Two curves are equivalent if they represent two motions with the same velocity at this instant.

Exercises

1. Show that the map $[x(t)] \to (\dot{x}^1(0), \ldots, \dot{x}^n(0))$ is one-to-one and onto V^n, so that it can be used to define the structure of a vector space on the collection $T_a(\boldsymbol{R}^n)$ of equivalence classes.
2. Prove that this definition of $T_a(\boldsymbol{R}^n)$ is equivalent to the earlier one of the present section.
3. Using a standard method of definition of the tangent plane to S^2, the unit sphere in $\boldsymbol{R}^3$, show that the vectors of $T_a(\boldsymbol{R}^3)$, $a \in S^2$, which belong to equivalence classes $[x(t)]$ determined by curves lying on S^2, determine a subspace of $T_a(\boldsymbol{R}^3)$ and that this subspace may be naturally identified with the tangent plane to S^2 at a.
4. For each of the vectors E_{ia}, $a \in \boldsymbol{R}^n$ and $i = 1, \ldots, n$, identify the equivalence class of curves corresponding to it by defining a particularly simple curve in the class. This gives an interpretation of the canonical basis of $T_a(\boldsymbol{R}^n)$.

4 Another Definition of $T_a(\boldsymbol{R}^n)$

In this section we give a characterization of the space of tangent vectors attached to a point a of $\boldsymbol{R}^n$ which we shall later use in extending this concept to manifolds. In spite of its formal and abstract nature it is relatively easy to work with; it is hoped that some intuitive clarification has resulted from the earlier definitions.

Let us denote by $C^\infty(a)$ the collection of all C^∞ functions whose domain includes a, *identifying* those functions which agree on an open set containing a—since we are only interested in their derivatives at a. Let $X_a = \sum_{i=1}^n \alpha^i E_{ia}$ be the expression for a vector of $T_a(\boldsymbol{R}^n)$ in the canonical basis; we define the *directional derivative* Δf of f at a in the "direction of X_a" by $\Delta f = \sum_{i=1}^n \alpha^i \, \partial f/\partial x^i$, $\partial f/\partial x^i$ evaluated at $a = (a^1, \ldots, a^n)$. This is a slight extension of the usual definition in that we do not require X_a to be a unit vector. Since Δf depends on f, a, and X_a we shall write it as $X_a^* f$. Thus

$$X_a^* f = \sum_{i=1}^n \alpha^i \left(\frac{\partial f}{\partial x^i}\right)_a .$$

We may take the directional derivative in the "direction of X_a" of any C^∞ function defined in a neighborhood of a. Hence $f \to X_a^* f$ defines a mapping assigning to each $f \in C^\infty(a)$ a real number

$$X_a^* \colon C^\infty(a) \to \boldsymbol{R}.$$

It is reasonable to denote this mapping by $X_a^* = \sum_{i=1}^n \alpha_i(\partial/\partial x^i)$, where we must remember that the derivatives are to be evaluated at a. We remark that $X_a^* x^i = \alpha^i$, $i = 1, \ldots, n$, so that the vector X_a is completely determined if its

value on every C^∞ function at a is known—or even on the functions $f^i(x) = x^i$.

We have agreed not to distinguish between C^∞ functions f, g in $C^\infty(a)$ if they agree on some open set containing a. Two functions of $C^\infty(a)$ may be added or multiplied to give another element of $C^\infty(a)$, whose domain is the intersection of their domains. If $\alpha \in \boldsymbol{R}$, then αf is a C^∞ function with the same domain as f, so $f \in C^\infty(a)$ implies $\alpha f \in C^\infty(a)$; the same result would be obtained by multiplying f by a C^∞ function whose value is α on some open set about a. Thus $C^\infty(a)$ is an algebra over $\boldsymbol{R}$ containing $\boldsymbol{R}$ as a subalgebra. Remembering the fundamental properties of derivatives we see at once that if α, β are real numbers and f, g are C^∞ functions defined in open sets containing a, then we have

(i) $X_a^*(\alpha f + \beta g) = \alpha(X_a^* f) + \beta(X_a^* g)$ (linearity)

and

(ii) $X_a^*(fg) = (X_a^* f)g(a) + f(a)(X_a^* g)$ (Leibniz rule).

Let $\mathscr{D}(a)$ denote all mappings of $C^\infty(a)$ to $\boldsymbol{R}$ with these properties; we may call the elements of $\mathscr{D}(a)$ "*derivations*" on $C^\infty(a)$ *into* $\boldsymbol{R}$. We see that $\mathscr{D}(a)$ is a vector space over $\boldsymbol{R}$ for if $D_1, D_2 \colon C^\infty(a) \to \boldsymbol{R}$ and $\alpha, \beta \in \boldsymbol{R}$, then we *define* $(\alpha D_1 + \beta D_2)f = \alpha(D_1 f) + \beta(D_2 f)$, where the operations on the right are in $\boldsymbol{R}$. This defines in $\mathscr{D}(a)$ both addition and multiplication by real numbers α, β. This is the standard procedure for defining a vector space structure on maps of a set into a field. One must check that the vector space axioms are indeed satisfied by these operations. In particular, it must be verified that if $D \in \mathscr{D}(a)$, then $\alpha D \in \mathscr{D}(a)$, and if $D_1, D_2 \in \mathscr{D}(a)$, then so also are $D_1 + D_2$. This means checking the linearity of $\alpha D \colon C^\infty(a) \to \boldsymbol{R}$ and $D_1 + D_2 \colon C^\infty(a) \to \boldsymbol{R}$ and checking that the Leibniz rule is satisfied. We do this for γD only. Suppose then $\gamma, \alpha, \beta \in \boldsymbol{R}$, $D \in \mathscr{D}(a)$, and $f, g \in C^\infty(a)$. Then

$$\begin{aligned}
(\gamma D)(\alpha f + \beta g) &= \gamma[D(\alpha f + \beta g)] && \text{(by definition of } \gamma D\text{)}\\
&= \gamma[\alpha(Df) + \beta(Dg)] && \text{(by property (i))}\\
&= \gamma\alpha(Df) + \gamma\beta(Dg) && \text{(by the distributive law of } \boldsymbol{R}\text{)}\\
&= \alpha(\gamma D)f + \beta(\gamma D)g && \text{(by our definition of } \gamma D\text{).}
\end{aligned}$$

It follows that the map $\gamma D \colon C^\infty(a) \to \boldsymbol{R}$ is linear. That γD satisfies the Leibniz rule for differentiation of products is equally easy:

$$\begin{aligned}
(\gamma D)(fg) &= \gamma[D(fg)] && \text{(by definition of } \gamma D\text{)}\\
&= \gamma[(Df)g(a) + f(a)(Dg)] && \text{(by property (ii))}\\
&= \gamma(Df)g(a) + f(a)\gamma(Dg) && \text{(these being real numbers)}\\
&= ((\gamma D)f)g(a) + f(a)((\gamma D)g) && \text{(by definition of } \gamma D\text{).}
\end{aligned}$$

As was remarked, a similar verification shows that $D_1 + D_2$ is a derivation into $\boldsymbol{R}$; it is left as an exercise.

The correspondence $X_a \to X_a^*$ associates to each element X_a of $T_a(\boldsymbol{R}^n)$ an element of $\mathscr{D}(a)$, namely the mapping $X_a^*: C^\infty(a) \to \boldsymbol{R}$ defined by taking the directional derivative of $f \in C^\infty(a)$ at a in the direction X_a. This mapping from $T_a(\boldsymbol{R}^n) \to \mathscr{D}(a)$ is one-to-one since $X_a^* = Y_a^*$ means that $X_a^* f = Y_a^* f$ for every $f \in C^\infty(a)$ which implies $X_a = Y_a$. Indeed we have noted the ith component of X_a relative to the natural basis is just $X_a^* x^i$ so that $X_a = \sum_{i=1}^n (X_a^* x^i) E_{ia} = Y_a$. Finally, it is easy to see that this mapping is linear. If $Z_a = \alpha X_a + \beta Y_a \in T_a(\boldsymbol{R}^n)$, then for the directional derivatives we have for any $f \in C^\infty(a)$,

$$Z_a^* f = \alpha(X_a^* f) + \beta(Y_a^* f).$$

If interpreted in terms of the operations in $\mathscr{D}(a)$, this means exactly that the mapping $T_a(\boldsymbol{R}^n) \to \mathscr{D}(a)$ is linear. In summary then, $X_a \to X_a^*$ defines an isomorphism of the vector space $T_a(\boldsymbol{R}^n)$ into the vector space $\mathscr{D}(a)$, which allows us to identify $T_a(\boldsymbol{R}^n)$ with a subspace of $\mathscr{D}(a)$. However, more can be said; in fact this isomorphism is *onto*, and we have the following theorem.

(4.1) Theorem *The vector space $T_a(\boldsymbol{R}^n)$ is isomorphic to the vector space $\mathscr{D}(a)$ of all derivations of $C^\infty(a)$ into $\boldsymbol{R}$. This isomorphism is given by making each X_a correspond to the directional derivative X_a^* in the direction of X_a.*

To prove the theorem it only remains to show that every derivation of $C^\infty(a)$ into $\boldsymbol{R}$ is a directional derivative, that is, that $X_a \to X_a^*$ is a map *onto* $\mathscr{D}(a)$. This will result from two lemmas.

(4.2) Lemma *Let D be an arbitrary element of $\mathscr{D}(a)$. Then D is zero on any function $f \in C^\infty(a)$ which is constant in a neighborhood of a.*

Proof Because the map D is linear, it is enough to show that if 1 denotes the constant function of value 1, then $D1 = 0$. However, $D1 = D(1 \cdot 1) = (D1)1 + 1(D1) = D1 + D1 = 2D1$, so $D1 = 0$. We must remember in interpreting these equalities that multiplying $f \in C^\infty(a)$ by a real number α gives exactly the same result as multiplying by the C^∞ function whose value is constant and equal to α in some open set (possibly $\boldsymbol{R}^n$) containing a, at least as far as the algebra $C^\infty(a)$ is concerned: we have identified $\boldsymbol{R}$ with the subalgebra of such functions. ∎

(4.3) Lemma *Let $f(x^1, \ldots, x^n)$ be defined and C^∞ on some open set U. If $a \in U$, then there is a spherical neighborhood B of a, $B \subset U$, and C^∞-functions*

$g^1, \ldots, g^n$ defined on B such that:

(i) $g^i(a) = \left(\dfrac{\partial f}{\partial x^i}\right)_{x=a}$

and

(ii) $f(x^1, \ldots, x^n) = f(a) + \sum_{i=1}^n (x^i - a^i)g^i(x)$.

Proof Let $B \subset U$ be a spherical neighborhood of a and note that for $x \in B, f(x) = f(a) + \int_0^1 (\partial/\partial t)f(a + t(x - a))\, dt$. Hence,

$$f(x) = f(a) + \sum_{i=1}^n (x^i - a^i) \int_0^1 \left[\frac{\partial f}{\partial x^i}\right]_{a+t(x-a)} dt.$$

Let

$$g^i(x) = \int_0^1 \left[\frac{\partial f}{\partial x^i}\right]_{a+t(x-a)} dt, \qquad i = 1, \ldots, n;$$

these are C^∞-functions and satisfy the two conditions. ∎

Proof of Theorem 4.1 Using these lemmas we may complete the proof of Theorem 4.1. Suppose D is any derivation on $C^\infty(a)$. We wish to show that, given $D \in \mathscr{D}(a)$, there is a vector $X_a \in T_a(\mathbf{R}^n)$ such that for any $f \in C^\infty(a)$, we have $X_a^* f = Df$. If this be so, then $X_a^* = D$ and we see that every derivation of $C^\infty(a)$ into $\mathbf{R}$ is a directional derivative; thus the map $X_a \to X_a^*$ of $T_a(\mathbf{R}^n)$ to $\mathscr{D}(a)$ is an isomorphism onto.

Let $h^i(x^1, \ldots, x^n) = x^i$. Then denote by α^i the value of Dh^i, that is, $\alpha^i = Dh^i$. Consider $X_a = \sum_{i=1}^n \alpha_i E_{ia}$; as an operator on $C^\infty(a)$, it gives

$$X_a^* f = \sum_{i=1}^n \alpha^i \left(\frac{\partial f}{\partial x^i}\right)_a.$$

On the other hand, by Lemma 4.3, $f(x) = f(a) + \sum_{i=1}^n (x^i - a^i)g^i(x)$ on some $B_\varepsilon(a)$ in the domain of f. Restricting to $B_\varepsilon(a)$ and using the properties of D, we may write

$$Df = D(f(a)) + \sum_{i=1}^n \{(D(x^i - a^i))g^i(a) + 0 \cdot Dg^i\}.$$

By Lemma 4.2, $D(f(a)) = 0$ and $D(x^i - a^i) = Dx^i = \alpha^i$; and by Lemma 4.3, $g^i(a) = (\partial f/\partial x^i)_a$. Therefore $Df = \sum_{i=1}^n \alpha^i (\partial f/\partial x^i)_a = X_a^* f$. Since f is an arbitrary element of $C^\infty(a)$, we have $D = X_a^*$. This completes the proof. ∎

Theorem 4.1 allows us to *identify* the vector space $T_a(\mathbf{R}^n)$ with the space $\mathscr{D}(a)$ of linear operators on functions of $C^\infty(a)$ into $\mathbf{R}$ which satisfy the product rule of Leibniz, that is, the "*derivations into* $\mathbf{R}$."

Note that under this identification the canonical basis vectors $E_{1a}, \ldots, E_{na}$ of $T_a(\boldsymbol{R}^n)$ are identified with $\partial/\partial x^1, \ldots, \partial/\partial x^n$, the directional derivatives (evaluated at $x = a$) in the directions of the coordinate axes:

$$E_{ia} \to E_{ia}^* f = \left(\frac{\partial f}{\partial x^i}\right)_a .$$

We will make this identification from now on for vectors in $T_a(\boldsymbol{R}^n)$ and for this reason we will drop the asterisk $*$ which distinguishes the vector X_a as a segment or point pair from the directional derivative: $X_a^* f$ will be written $X_a f$. In $\boldsymbol{R}^n$ we may use either E_{ia} or $\partial/\partial x^i$ to denote the unit vector parallel to the ith coordinate axis. This characterization of $T_a(\boldsymbol{R}^n)$ requires C^∞ functions; although $C^r(a)$ is an algebra, it is known to have other derivations than directional derivatives. Our situation is then that we shall rely on Euclidean space for our geometric intuition of the space of tangent vectors at a point, but in formal definitions and proofs we will use the ideas above: a vector at a point is a linear operator of a certain kind—satisfying the product rule for derivatives—on the C^∞ functions at the point.

Exercises

1. Let $a \in \boldsymbol{R}^n$ and f, g be two C^∞ functions whose domains of definition both contain a. We shall say f is *equivalent* to g at a, $f \sim g$, if and only if they agree on some open set containing a. Show that this is an equivalence relation and that the collection of equivalence classes, which we call *germs* of C^∞ functions at a, is an algebra over $\boldsymbol{R}$. [It is precisely *this* algebra which is $C^\infty(a)$.]
2. Do the statements of Exercise 1 remain true if we replace C^∞ by C^r, $r \geq 0$? Suppose that we merely require equality of f and g and all of their derivatives at a, does this give an equivalence relation and an algebra over $\boldsymbol{R}$, or is equality on an open set needed?
3. Prove that the collection of all maps of a set X into a field F has a natural vector space structure. Follow the definitions indicated for $\mathscr{D}(a)$.
4. Show that if D_1 and D_2 are in $\mathscr{D}(a)$, then $D_1 + D_2$ is also in $\mathscr{D}(a)$.
5. Using the definition (Section 3) of equivalence of C^1 curves through $a \in \boldsymbol{R}^n$, prove that the mapping of $T_a(\boldsymbol{R}^n)$, defined as "velocity vectors," to $\mathscr{D}(a)$, taking the class $[x(t)]$ to the operator D defined by

$$Df = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i}\right)_a \dot{x}^i(0) ,$$

is independent of the choice of the curve $x(t)$ in $[x(t)]$ and determines an isomorphism of $T_a(\boldsymbol{R}^n)$ onto $\mathscr{D}(a)$.

5 Vector Fields on Open Subsets of R^n

A *vector field* on an open subset $U \subset \boldsymbol{R}^n$ is a function which assigns to each point $p \in U$ a vector $X_p \in T_p(\boldsymbol{R}^n)$. A similar definition applies to Euclidean space $\boldsymbol{E}^n$. There are many examples in physics for $n = 2$ and $n = 3$. The best known is the gravitational field: If an object of mass μ is located at a point 0, then to each point p in $U = \boldsymbol{E}^n - \{0\}$, there is assigned a vector which denotes the force of attraction on a particle of unit mass placed at the point. This vector is represented by a line segment or arrow from p (as initial point) directed toward 0 and having length $k\mu/r^2$, r denoting the distance $d(0, p)$ and k a fixed constant determined by the units chosen (see

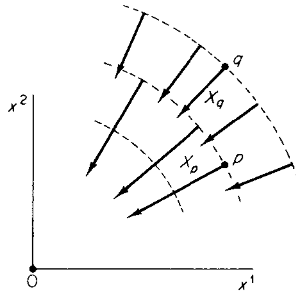


Figure II.4
First quadrant portion of gravitational field of point mass at origin.

Fig. II.4). If we introduce Cartesian coordinates with 0 as origin and $\mu = 1$, then for the point p with coordinates (x^1, x^2, x^3) the components of X_p in the canonical basis are

$$\frac{-kx^1}{r^3}, \frac{-kx^2}{r^3}, \frac{-kx^3}{r^3} \qquad \text{with} \quad r = ((x^1)^2 + (x^2)^2 + (x^3)^2)^{1/2},$$

that is,

$$X_p = \frac{-k}{r^3}(x^1 E_{ip} + x^2 E_{2p} + x^3 E_{3p}) = \frac{-k}{r^3}\left(x^1 \frac{\partial}{\partial x^1} + x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3}\right).$$

We note that the components of X_p are C^∞ functions of the coordinates. We shall say that a vector field on $\boldsymbol{R}^n$ is C^∞ or *smooth* if its components relative to the canonical basis are C^∞ functions on U. Unless otherwise stated, all vector fields considered will be assumed to have this property, although it is quite possible to define continuous, C^1, and so on, vector fields also. When dealing with vector fields, as with functions, the independent variable will be omitted from the notation. Thus we write X rather than X_p

just as we customarily use f rather than $f(p)$ for a function. Then X_p is the *value at* p of X, that is, the vector of the field which is attached to p—it lies in $T_p(\boldsymbol{R}^n)$.

Further examples of vector fields are given for each $i = 1, \ldots, n$ by the fields $E_i = \partial/\partial x^i$ which assign to every $p \in \boldsymbol{R}^n$ the naturally defined basis vector E_i at that point. The vector fields $E_1, \ldots, E_n$ being independent, even orthogonal unit vectors, at each point p form a basis there of $T_p(\boldsymbol{R}^n)$; such a set of fields is called a *field of frames*. The vector fields X_1, X_2 on $U = \boldsymbol{R}^2 - \{0\}$ defined by $X_1 = x^1E_1 + x^2E_2$ and $X_2 = x^2E_1 - x^1E_2$ also define a field of frames; geometrically X_{1p} is a vector along a ray from 0 to p and X_{2p} is a vector perpendicular to it, that is, tangent to the circle through p with center at 0 (see Fig. II.5). It is often convenient, as we know from elementary mechanics to use other frames (even in Euclidean space) than E_1 and E_2.

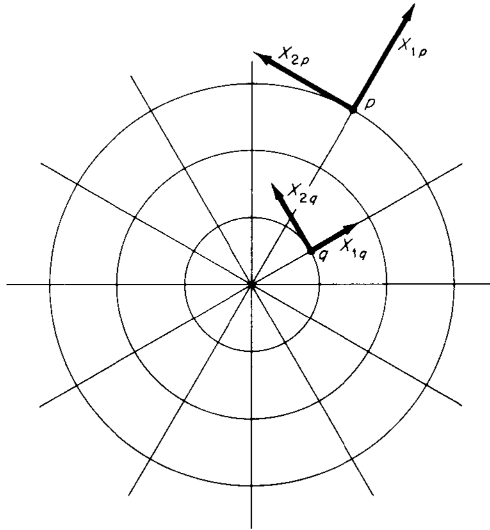


Figure II.5

If X is a C^∞-vector field on U and f a C^∞ function on U, then Xf is the C^∞-function on U defined by $(Xf)(p) = X_pf$. Indeed, if the components of X are the functions $\alpha^1(p), \ldots, \alpha^n(p)$ so that $X = \sum_{i=1}^n \alpha^i E_i$, then

$$(Xf)(p) = \sum_{i=1}^n \alpha^i(p)\left(\frac{\partial f}{\partial x^i}\right)_p.$$

We see from the right-hand side that Xf is a C^∞ function of p on U since $\alpha^i(p) \in C^\infty(U)$ and $\partial f/\partial x^i \in C^\infty(U)$. Thus $f \mapsto Xf$ maps $C^\infty(U) \to C^\infty(U)$.

We note also that $C^\infty(U)$ is an algebra over R with unit, where $\boldsymbol{R}$ is identified with the constant functions and, in particular, the constant function 1 with the unit. It is natural to ask whether X is a linear map of $C^\infty(U)$ to $C^\infty(U)$ and more generally whether it is a *derivation*, that is, satisfies the Leibniz product rule. In fact, this is so, for we may write

$$\begin{aligned}[X(\alpha f + \beta g)](p) &= X_p(\alpha f + \beta g) = \alpha(X_p f) + \beta(X_p g)\\ &= \alpha(Xf)(p) + \beta(Xg)(p),\end{aligned}$$

and

$$\begin{aligned}[X(fg)](p) &= (X_p f)g(p) + f(p)(X_p g)\\ &= [(Xf)(p)]g(p) + f(p)[(Xg)(p)].\end{aligned}$$

Since the functions on the right and left agree for each $p \in U$, they are equal as functions. Thus $X\colon C^\infty(U) \to C^\infty(U)$ is a derivation which maps $C^\infty(U)$ into itself, a slight variation from the previous case. (This, in fact, is the customary use of the term "derivation" of an algebra. If A is an algebra over $\boldsymbol{R}$, then a *derivation* is a map $D\colon A \to A$ which is linear and satisfies the product rule of Leibniz. For example, $\partial/\partial x$ is a derivation on the algebra of all polynomials in two variables x and y.)

We conclude this section by proving an important property of C^∞-functions which, with the corollary given here, is used very often in discussions of vector fields (see the exercises). It is a "separation theorem" and contrasts strongly the behavior of C^∞ and C^ω functions on $\boldsymbol{R}^n$. (There exist stronger versions of this theorem as we shall see later.)

(5.1) Theorem *Let $F \subset \boldsymbol{R}^n$ be a closed set and $K \subset \boldsymbol{R}^n$ compact, $F \cap K = \varnothing$. Then there is a C^∞ function $\sigma(x)$ whose domain is all of $\boldsymbol{R}^n$ and whose range of values is the closed interval* $[0, 1]$ *such that $\sigma(x) \equiv 1$ on K and $\sigma(x) \equiv 0$ on F.*

Proof We prove the theorem in two steps.

(a) Let $B_\varepsilon(a)$ be an open ball of center a and radius ε. We show that there is a C^∞ function $g(x)$ on $\boldsymbol{R}^n$ which is positive on $B_\varepsilon(a)$, identically 1 on $\bar{B}_{\varepsilon/2}$, and 0 outside $B_\varepsilon(a)$. The function $h(t)$ defined by $h(0) = 0$ for $t \leq 0$ and $h(t) = e^{-1/t}$ for $t > 0$ is C^∞ since we can prove by direct computation that all of its derivatives exist and are zero at $t = 0$, and since it is analytic for other values of t. We let

$$\bar{g}(x) = \frac{h(\varepsilon^2 - \|x\|^2)}{h(\varepsilon^2 - \|x\|^2) + h(\|x\|^2 - \frac{1}{4}\varepsilon^2)}.$$

Since the denominator is never zero [at either $\varepsilon^2 - \|x\|^2$ or $\|x\|^2 - \frac{1}{4}\varepsilon^2$ or at both h is positive], this is a C^∞ function. When $\|x\| \geq \varepsilon$, the numerator is

zero, otherwise it is positive; and when $0 \le \|x\| \le \frac{1}{2}\varepsilon$, the value of $\bar{g}(x)$ is identically 1. Thus $\bar{g}(x)$ is C^∞, vanishes outside $B_\varepsilon(0)$, and is positive on its interior; in fact $\bar{g}(x) = 1$ for $x \in \bar{B}_{\varepsilon/2}(0)$. Hence $g(x) = \bar{g}(x - a)$ has the desired properties.

(b) For step two let $B_\varepsilon(a_i)$, $i = 1, \ldots, k$, be a finite collection of n-balls in $\boldsymbol{R}^n - F$ such that $\bigcup_{i=1}^k B_{\varepsilon/2}(a_i) \supset K$. It is possible to find such a collection because of the compactness of K and since $K \cap F = \varnothing$. For $B_\varepsilon(a_i)$ let $g_i(x)$ have the properties above and define $\sigma(x)$ by

$$\sigma(x) = 1 - \prod_{i=1}^{k} (1 - g_i).$$

On each $x \in K$ at least one g_i has the value 1, $\sigma(x) = 1$, so $\sigma \equiv 1$ on K. Outside $\bigcup_{i=1}^k B_\varepsilon(a_i)$ each g_i vanishes so $\sigma(x) = 0$ and, since F lies outside this union, $\sigma \equiv 0$ on F. This completes the proof. ∎

(5.2) Corollary *Let $f(x^1, \ldots, x^n)$ be C^∞ on an open set $U \subset \boldsymbol{R}^n$ and let $a \in U$. Then there is an open set $V \subset U$, which is a neighborhood of a, and a C^∞ function $f^*(x^1, \ldots, x^n)$ defined on all of $\boldsymbol{R}^n$ such that $f^*(x) = f(x)$ for all $x \in V$ and $f^*(x) = 0$ for x outside U.*

Proof Choose any neighborhoods V_1, V_2 of a such that $\bar{V}_1 \subset V_2$, $\bar{V}_2 \subset U$ and $\bar{V}_1$ is compact. Let $\bar{V}_1 = K$ and $F = \boldsymbol{R}^n - V_2$ in Theorem 5.1. Then take $\sigma(x)$, a C^∞ function whose value is 1 on $\bar{V}_1$ and 0 outside V_2, that is, on F. Define $f^*(x) = \sigma(x) f(x)$ for $x \in U$ and $f^*(x) = 0$ for $x \in \boldsymbol{R}^n - V_2$. Since f^* thus defined is C^∞ on U, where it is equal to σf, and is C^∞ on $\boldsymbol{R}^n - \bar{V}_2$ where it is identically zero, and since on the (open) intersection $U - \bar{V}_2$ of these two sets, both definitions agree, we see that f^* is C^∞ on $\boldsymbol{R}^n$ and has the properties needed. ∎

Exercises

1. Prove that the function $h(t)$ defined above does in fact have derivatives of all orders at $t = 0$ and that they all have the value zero there. Why does this imply that $h(t)$ is not analytic at $t = 0$?
2. Suppose a vector $X_p \in T_p(\boldsymbol{R}^n)$ is given at each $p \in U$, an open subset of $\boldsymbol{R}^n$. Show that this defines a C^∞ vector field if and only if for each $f \in C^\infty(U)$, Xf is C^∞ on U.
3. Let D be a derivation on the algebra $C^\infty(U)$. If f, g are in $C^\infty(U)$ and $f(x) = g(x)$ for all $x \in V$, an open subset of U, then prove $(Df)(x) = (Dg)(x)$ at each point $x \in V$.
4. Using the preceding problem show that the derivations $\mathscr{D}(U)$ on $C^\infty(U)$ define a vector space over $\boldsymbol{R}$ and that this space is isomorphic with the space of C^∞ vector fields on U, which we denote $\mathfrak{X}(U)$. Show that these are infinite-dimensional vector spaces.

5. If D_1 and D_2 are derivations on an algebra A, show by example that $D_1 D_2$ need not be a derivation but that $D_1 D_2 - D_2 D_1$ is a derivation. If we consider $A = C^\infty(U)$, U open $\subset \mathbf{R}^n$, $X = \sum \alpha^i(x)\, \partial/\partial x^i$ and $Y = \sum \beta^i(x)\, \partial/\partial y^i \in \mathfrak{X}(U)$, then this means that $XY - YX = \sum \gamma^i(x)\, \partial/\partial x^i$. Verify this and compute $\gamma_i(x)$.
6. Give examples of linear operators on $C^\infty(U)$, U an open subset of $\mathbf{R}^n$, which are *not* derivations of $C^\infty(U)$.
7. Let $p \in U$, an open subset of $\mathbf{R}^n$, and let $X_p \in T_p(\mathbf{R}^n)$ be a vector at p. Show that X_p may be extended to a C^∞ vector field X on U.
8. In Theorem 5.1, assume only that K is closed (not necessarily compact). Does the theorem still hold?

6 The Inverse Function Theorem

In order to simplify the terminology of this and later sections we introduce the notion of diffeomorphism, or differentiable homeomorphism, between two spaces. Of course this concept can have no meaning unless the spaces are such that differentiability is defined, which at the present moment means that they must be subsets of Euclidean spaces. Therefore without prejudice to a more general later treatment, let us suppose that $U \subset \mathbf{R}^n$ and $V \subset \mathbf{R}^n$ are open sets. We then shall say that a mapping $F\colon U \to V$ is a C^r*-diffeomorphism* if: (i) F is a homeomorphism and (ii) *both* F and F^{-1} are of class C^r, $r \geq 1$ (when $r = \infty$ we simply say *diffeomorphism*). It is perhaps not obvious why we need to require both F and F^{-1} to be of class C^r—it is because we wish the relation to be symmetric. As the following example shows, the differentiability of F^{-1} is not a consequence of that of F, even when F is a homeomorphism. Let $U = \mathbf{R}$ and $V = \mathbf{R}$ and $F\colon t \mapsto s = t^3$; this is a homeomorphism and F is analytic but $F^{-1}\colon s \mapsto t = s^{1/3}$ is not C^1 on V since it has no derivative at $s = 0$.

One might suspect that our definition contains some redundant requirements as in fact it does—in two ways. First, as shown in Exercise 6, it would not be possible to have a diffeomorphism between open subspaces of Euclidean spaces of different dimensions; indeed a famous theorem of algebraic topology (Brouwer's invariance of domain) asserts that even a *homeomorphism* between open subsets of $\mathbf{R}^n$ and $\mathbf{R}^m$, $m \neq n$, is impossible. Secondly, in the example given above the derivative of F vanishes at $t = 0$, thus behaving atypically: If it vanished everywhere, then F could not be a homeomorphism of $\mathbf{R}$ to $\mathbf{R}$ and if it vanished at no point, then F^{-1} would indeed be a differentiable map (please argue this through!). We can certainly see at once that if $F\colon U \to V$ is a homeomorphism and both F and F^{-1} are of class C^1 at least, then $DF(x)$ is nonsingular, that is, has nonvanishing determinant at each $x \in U$; for $F^{-1} \circ F = I$, the identity map of U to U and by the chain rule $DI(x) = DF^{-1}(F(x)) \cdot DF(x)$. Since $DI(x)$ is just the identity

matrix for every $x \in U$, $DF(x)$ is nonsingular and its determinant is never zero. This includes the assertion that, if $F: \boldsymbol{R} \to \boldsymbol{R}$ is a diffeomorphism, then its derivative can never be zero. The main theorem of this section will have as a consequence Corollary 6.7, which is the converse of this statement. Before proving it, however, we consider two examples of diffeomorphisms of $\boldsymbol{R}^n$ to $\boldsymbol{R}^n$.

(6.1) Example Let $F: \boldsymbol{R}^n \to \boldsymbol{R}^n$ be the translation taking $a = (a^1, \ldots, a^n)$ to $b = (b^1, \ldots, b^n)$. Then F is given by

$$F(x^1, \ldots, x^n) = (x^1 + (b^1 - a^1), \ldots, x^n + (b^n - a^n)),$$

or $F(x) = x + (b - a)$. The coordinate functions $f^i(x) = x^i + (b^i - a^i)$ are analytic, and hence C^∞. The translation $G(x) = x + (a - b)$ is F^{-1} which is then also C^∞ and since F, F^{-1} are defined and continuous, F is a homeomorphism. Thus F is a diffeomorphism.

(6.2) Example Let $F: \boldsymbol{R}^n \to \boldsymbol{R}^n$ be a linear transformation

$$F(x^1, \ldots, x^n) = \left(\sum_{j=1}^{n} \alpha_j^1 x^j, \ldots, \sum_{j=1}^{n} \alpha_j^n x^j\right),$$

or, using matrix notation with x as an $n \times 1$ (column matrix) and $A = (\alpha_j^i)$, $F(x) = Ax$. Computation shows that $DF(x)$ is the constant matrix A, $DF(x) = A$. If $\det A \neq 0$, then A has an inverse B and the homogeneous linear transformation $G(x) = Bx$ is F^{-1}. On the other hand if $\det A = 0$, then F is not one-to-one, in fact it maps at least a line through the origin onto the single point $0 = (0, 0, \ldots, 0)$. Obviously F is analytic and C^∞ in either case, so that F is a diffeomorphism if and only if $DF(x) = A$ is nonsingular.

Diffeomorphism is an equivalence relation among the open subsets of $\boldsymbol{R}^n$. We have the following lemma, whose proof we leave as an exercise, which gives the transitivity property; symmetry and reflexivity are part of the definition.

(6.3) Lemma *Let U, V, W be open subsets of $\boldsymbol{R}^n$, $F: U \to V$, $G: V \to W$ mappings onto, and $H = G \circ F: U \to W$ their composition. If any two of these maps is a diffeomorphism, then the third is also.*

We now state the main theorem of the section.

(6.4) Theorem (Inverse Function Theorem) *Let W be an open subset of $\boldsymbol{R}^n$ and $F: W \to \boldsymbol{R}^n$ a C^r mapping, $r = 1, 2, \ldots$, or ∞. If $a \in W$ and $DF(a)$ is nonsingular, then there exists an open neighborhood U of a in W such that*

$V = F(U)$ is open and $F: U \to V$ is a C^r diffeomorphism. If $x \in U$ and $y = F(x)$, then we have the following formula for the derivatives of F^{-1} at y:

$$DF^{-1}(y) = (DF(x))^{-1},$$

the term on the right denoting the inverse matrix to $DF(x)$.

This is one of the two basic theorems of analysis on which all of the theory in this book depends; the other is the existence theorem for ordinary differential equations (Chapter IV). The proof used here depends on the following fixed point theorem; a variety of proofs may be found in advanced calculus books.

(6.5) Theorem (Contracting Mapping Theorem) *Let M be a complete metric space with metric $d(x, y)$ and let $T: M \to M$ be a mapping of M into itself. Assume that there is a constant λ, $0 \leq \lambda < 1$, such that for all $x, y \in M$,*

$$d(T(x), T(y)) \leq \lambda\, d(x, y).$$

Then T has a unique fixed point a in M.

Proof Applying T repeatedly we see that $d(T^n(x), T^n(y)) \leq \lambda^n\, d(x, y)$. In particular, if we choose arbitrarily $x_0 \in M$ and let $x_n = T^n(x_0)$, then we assert that $d(x_n, x_{n+m}) \leq \lambda^n K$, $K \geq 0$, a constant independent of n, m. Using $T^{n+m}(x_0) = T^n(T^m(x_0))$, we write $d(x_n, x_{n+m}) \leq \lambda^n\, d(x_0, T^m(x_0))$. By the triangle inequality

$$\begin{aligned} d(x_0, T^m(x_0)) &\leq d(x_0, T(x_0)) + d(T(x_0), T^2(x_0)) \\ &\quad + \cdots + d(T^{m-1}(x_0), T^m(x_0)) \\ &\leq (1 + \lambda + \lambda^2 + \cdots + \lambda^{m-1})\, d(x_0, T(x_0)) \leq \frac{1}{1-\lambda} d(x_0, T(x_0)). \end{aligned}$$

This shows that we may take

$$K = \frac{1}{1-\lambda} d(x_0, T(x_0))$$

and proves the assertion. Thus $\{x_n\}$ is a Cauchy sequence and has a limit point a. Since $T(x_n) = x_{n+1}$ obviously has the same limit we see that

$$d(T(a), a) = \lim d(T(x_n), x_n) = \lim d(x_{n+1}, x_n) = 0$$

so $T(a) = a$ and a is a fixed point of T. There could not be two fixed points a, b for then $d(a, b) = d(T(a), T(b)) \leq \lambda\, d(a, b)$ contradicting the fact that $\lambda < 1$. ∎

Proof of Theorem 6.4 We shall organize the proof of the inverse function theorem in several steps in order to make it somewhat easier to follow.

(i) *We assume $F(0) = 0$ and $DF(0) = I$, the identity matrix.*

This may be done without loss of generality by virtue of Lemma 6.3 combined with the use of Examples 6.1 and 6.2. Next we define the mapping G on the same domain by

$$G(x) = x - F(x).$$

Then, obviously, $G(0) = 0$ and $DG(0) = 0$. (*Note*: In the last equation the right-hand side is the 0 matrix.)

(ii) *There exists a real number $r > 0$ such that DF is nonsingular on the closed ball $\bar{B}_{2r}(0) \subset W$ and for $x_1, x_2 \in \bar{B}_r(0)$ we have*

$$(*) \qquad \|G(x_1) - G(x_2)\| \leq \tfrac{1}{2}\|x_1 - x_2\|$$

and

$$(**) \qquad \|x_1 - x_2\| \leq 2\|F(x_1) - F(x_2)\|.$$

To verify these statements we choose r so that $\bar{B}_{2r}(0) \subset W$; further so that $\det(DF(x))$, which is a continuous function of x and not zero at 0, does not vanish on $\bar{B}_{2r}(0)$; and finally so that the derivatives of the coordinate functions of G, all of which are zero at 0, are bounded in absolute value by $1/2n$ on $\bar{B}_{2r}(0)$. With these assumptions, $x_1, x_2 \in \bar{B}_r(0)$ implies $\|x_1 - x_2\| \leq 2r$ and Theorem 2.2 with $m = n$ gives $(*)$. Inequality $(**)$ results from replacing $G(x_i)$ by $x_i - F(x_i)$, $i = 1, 2$, in $(*)$ and using a standard property of norms:

$$\|x_1 - F(x_1) - x_2 + F(x_2)\| \leq \tfrac{1}{2}\|x_1 - x_2\|$$

by $(*)$, but

$$\|x_1 - x_2\| - \|F(x_1) - F(x_2)\| \leq \|(x_1 - x_2) + F(x_1) - F(x_2)\|.$$

Combining these gives $(**)$. As a consequence of $(*)$ we obtain the following:

(iii) *If $\|x\| \leq r$, then $\|G(x)\| \leq r/2$, that is, $G(\bar{B}_r(0)) \subset \bar{B}_{r/2}(0)$. Moreover for each $y \in \bar{B}_{r/2}(0)$ there exists a unique $x \in \bar{B}_r(0)$ such that $F(x) = y$.*

The first statement is immediately obtained from $(*)$ by setting $x_1 = x$ and $x_2 = 0$; the second uses Lemma 6.3. If $y \in \bar{B}_{r/2}(0)$ and $x \in \bar{B}_r(0)$, then

$$\|y + G(x)\| \leq \|y\| + \|G(x)\| \leq \tfrac{1}{2}r + \tfrac{1}{2}r = r.$$

Let a mapping $T_y\colon \bar{B}_r(0) \to \bar{B}_r(0)$ be defined for $y \in \bar{B}_{r/2}(0)$ by $T_y(x) = y + G(x)$. Then $T_y(x) = x$ if and only if $y = x - G(x)$ or, equivalently, $F(x) = y$. However, inequality $(*)$ in the form

$$\|T_y(x_1) - T_y(x_2)\| = \|G(x_1) - G(x_2)\| \leq \tfrac{1}{2}\|x_1 - x_2\|,$$

valid for $x_1, x_2 \in \bar{B}_r(0)$, implies that $T_y(x)$ is a contracting mapping of the compact set $\bar{B}_r(0)$ into itself. Therefore by Theorem 6.5 there is a unique x such that $y = F(x)$. Since this is valid for any $y \in \bar{B}_{r/2}(0)$, we see that F^{-1} is defined on that set. In particular, F being continuous, $U = F^{-1}(B_{r/2}(0))$ is an open subset of $\bar{B}_r(0)$. Let $V = B_{r/2}(0)$; since $\bar{B}_r(0) \subset W$ we see that:

(iv) *F is a homeomorphism of the open set $U \subset W$ onto the open set V.*

It remains only to prove continuity of F^{-1}, which is a consequence of the inequality (∗∗). Whenever $x_1, x_2 \in U$, we have $y_1 = F(x_1)$ and $y_2 = F(x_2)$, and (∗∗) becomes

$$\|F^{-1}(y_1) - F^{-1}(y_2)\| \leq 2\|y_1 - y_2\|,$$

which implies that $F^{-1}\colon V \to U$ is continuous.

(v) *Let $b = F(a)$ be in V. Then F^{-1} is differentiable at b and $DF^{-1}(b) = [DF(a)]^{-1}$, the matrix inverse to $DF(a)$.*

Since F is of class C^r, $r \geq 1$, on W it is differentiable on all of U, in particular, at $a = F^{-1}(b)$. Thus by definition

$$F(x) - F(a) = DF(a) \cdot (x - a) + \|x - a\| r(x, a),$$

where $r(x, a) \to 0$ as $x \to a$. By (ii), $DF(a)$ is nonsingular and we let A be its inverse matrix. Multiplying the above expression by A and using $y = F(x)$, $x = F^{-1}(y)$, and $a = F^{-1}(b)$, and so on, we obtain

$$\begin{aligned} A \cdot (y - b) = {} & F^{-1}(y) - F^{-1}(b) \\ & + \|F^{-1}(y) - F^{-1}(b)\| A \cdot r(F^{-1}(y), F^{-1}(b)). \end{aligned}$$

This, in turn, gives

$$F^{-1}(y) = F^{-1}(b) + A \cdot (y - b) + \|y - b\| \tilde{r}(y, b)$$

if we suppose $y \neq b$ and define

$$\tilde{r}(y, b) = -\frac{\|F^{-1}(y) - F^{-1}(b)\|}{\|y - b\|} A \cdot r(F^{-1}(y), F^{-1}(b)).$$

Inequality (∗∗) shows that the initial fraction is bounded by 2, A is a matrix of constants, and $F^{-1}(y)$ is continuous so it is clear that $\lim_{y \to b} \tilde{r}(y, b) = 0$ which proves the differentiability of F^{-1} at any $b \in V$ and shows that

$$DF^{-1}(b) = A = [DF(a)]^{-1}$$

as claimed. The following statement completes the proof.

(vi) *If F is of class C^r on U, then F^{-1} is of class C^r on V.*

For $y \in V$ we have just seen that

$$DF^{-1}(y) = [DF(F^{-1}(y))]^{-1}.$$

Since $F^{-1}(y)$ is continuous as a function of y on V and its range is U, since DF is of class C^{r-1} and nonsingular on U, and since, finally, the entries in the inverse of a nonsingular matrix are C^{∞} functions of the entries of the matrix, it follows that DF^{-1} is continuous on V, thus F^{-1} is of class C^1 at least. In fact if F^{-1} is of class $k < r$, the entries of DF^{-1} are of class $k - 1$ at least on V, but the formula above for them shows these entries to be given by composition of functions of class C^k or greater and hence to be of class C^k at least. This implies F^{-1} is of class C^{k+1}; so by induction F^{-1} is of class C^r. This completes the proof. ∎

The following two corollaries are immediate consequences of Theorem 6.4. We use the notation of the theorem, that is, W is an open subset of $\boldsymbol{R}^n$ and $F: W \to \boldsymbol{R}^n$.

(6.6) Corollary *If DF is nonsingular at every point of W, then F is an open mapping of W, that is, it carries W and open subsets of $\boldsymbol{R}^n$ contained in W to open subsets of $\boldsymbol{R}^n$.*

(6.7) Corollary *A necessary and sufficient condition for the C^{∞} map F to be a diffeomorphism from W to $F(W)$ is that it be one-to-one and DF be nonsingular at every point of W.*

Exercises

1. Carry the proof of Theorem 6.4 through in detail for the case $r = 1$, making any simplifications you can.
2. Prove Lemma 6.3.
3. Compute the Jacobian matrix for Examples 6.1 and 6.2.
4. Show that for transformations on a compact subset K of $\boldsymbol{R}^n$ which satisfy the conditions of the contracting mapping theorem except that $0 \leq \lambda \leq 1$ (we allow $\lambda = 1$), there still exists a fixed point. [*Hint:* Consider mappings $T_n = ((n-1)/n)T$.]
5. Prove Corollary 6.6.
6. Prove Corollary 6.7.
7. Prove that there does not exist a C^1 diffeomorphism from an open subset of $\boldsymbol{R}^n$ to an open subset of $\boldsymbol{R}^m$ if $m < n$.

7 The Rank of a Mapping

In linear algebra the *rank* of an $m \times n$ matrix A is defined in three equivalent ways: (i) the dimension of the subspace of V^n spanned by the rows, (ii) the dimension of the subspace of V^m spanned by the columns, and

(iii) the maximum order of any nonvanishing minor determinant. We see at once from (i) and (ii) that the rank $A \leq m, n$. The *rank* of a linear transformation is defined to be the dimension of the image, and one proves that this is the rank of any matrix which represents the transformation. From this it follows that, if P and Q are nonsingular matrices, then $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$.

When $F\colon U \to \boldsymbol{R}^m$ is a C^1 mapping of an open set $U \subset \boldsymbol{R}^n$, then $DF(x)$ has a rank at each $x \in U$. Because the value of a determinant is a continuous function of its entries, we see from (iii) that if rank $DF(a) = k$, then for some open neighborhood V of a, rank $DF(x) \geq k$; and, if $k = \inf(m, n)$, then rank $DF(x) = k$ on this V. In general, the inequality is possible:

$$F(x^1, x^2) = ((x^1)^2 + (x^2)^2, 2x^1x^2)$$

has Jacobian

$$DF(x^1, x^2) = \begin{pmatrix} 2x^1 & 2x^2 \\ 2x^2 & 2x^1 \end{pmatrix},$$

whose rank is 2 on all of $\boldsymbol{R}^2$ except the lines $x^2 = \pm x^1$. The rank is 1 on these lines except at $(0, 0)$ where it is zero.

We shall refer to the rank of $DF(x)$ as the *rank of F at x*. If we compose F with diffeomorphisms, then the facts cited and the chain rule imply that the rank of the composition is the rank of F, since diffeomorphisms have nonsingular Jacobians. We say F has rank k on a set A, if it has rank k for each $x \in A$. We use these definitions in stating the following basic theorem.

(7.1) Theorem (Rank Theorem) *Let $A_0 \subset \boldsymbol{R}^n$, $B_0 \subset \boldsymbol{R}^m$ be open sets, $F\colon A_0 \to B_0$ be a C^r mapping, and suppose the rank of F on A_0 to be equal to k. If $a \in A_0$ and $b = F(a)$, then there exist open sets $A \subset A_0$ and $B \subset B_0$ with $a \in A$ and $b \in B$, and there exist C^r diffeomorphisms $G\colon A \to U$ (open) $\subset \boldsymbol{R}^n$, $H\colon B \to V$ (open) $\subset \boldsymbol{R}^m$ such that $H \circ F \circ G^{-1}(U) \subset V$ and such that this map has the simple form*

$$H \circ F \circ G^{-1}(x^1, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0).$$

Before proceeding to the proof we make some general comments. This is clearly an important theorem for it tells us that a mapping of constant rank k behaves *locally* like projection of $\boldsymbol{R}^n = \boldsymbol{R}^k \times \boldsymbol{R}^{n-k}$ to $\boldsymbol{R}^k$ followed by injection of $\boldsymbol{R}^k$ onto $\boldsymbol{R}^k \times \{0\} \subset \boldsymbol{R}^k \times \boldsymbol{R}^{m-k} = \boldsymbol{R}^m$. This is an important tool and we shall use it frequently; later it will be rephrased in terms of local coordinates (cf. Chapter III, Remark 4.2). It implies Theorem 6.4 as a special case.

Proof To begin with, we may suppose $a = 0$, the origin of $\boldsymbol{R}^n$, and $b = 0$, the origin of $\boldsymbol{R}^m$. If the theorem holds for this case, then it may be seen to hold in general since composition of F with two translations gives $\tilde{F}(u) = F(u + a) - b$, which has the property that $\tilde{F}(0) = 0$. By similar

suppose that a $k \times k$ minor of nonzero determinant in $DF(a)$ is

$$\frac{\partial(f^1,\ldots,f^k)}{\partial(u^1,\ldots,u^k)} = \begin{pmatrix} \dfrac{\partial f^1}{\partial u^1} & \cdots & \dfrac{\partial f^1}{\partial u^k} \\ \vdots & & \vdots \\ \dfrac{\partial f^k}{\partial u^1} & \cdots & \dfrac{\partial f^k}{\partial u^k} \end{pmatrix}_{u=a},$$

the upper left $k \times k$ minor.

We define the C^r-mapping $G\colon A_0 \to \boldsymbol{R}^n$ by

$$G(u^1,\ldots,u^n) = (f^1(u^1,\ldots,u^n),\ldots,f^k(u^1,\ldots,u^n), u^{k+1},\ldots,u^n).$$

Then

$$DG = \left(\begin{array}{ccc|c} \dfrac{\partial f^1}{\partial u^1} & \cdots & \dfrac{\partial f^1}{\partial u^k} & \\ \vdots & & \vdots & * \\ \dfrac{\partial f^k}{\partial u^1} & \cdots & \dfrac{\partial f^k}{\partial u^n} & \\ \hline & 0 & & I_{n-k} \end{array}\right),$$

where I_{n-k} is the $(n-k) \times (n-k)$ identity, the terms in the lower left block are zero, and those in the upper right do not interest us. This matrix is nonsingular at $u = a$, hence there is in A_0 an open set A_1 containing a on which G is a diffeomorphism onto an open subset $U_1 = G(A_1)$. From the expression for G and the definition of U_1 we have $F \circ G^{-1}(0) = 0$, $F \circ G^{-1}(U_1) \subset B_0$, and

$$F \circ G^{-1}(x^1,\ldots,x^k,x^{k+1},\ldots,x^n) = (x^1,\ldots,x^k,\bar{f}^{k+1}(x),\ldots,\bar{f}^m(x))$$

with $\bar{f}^{k+j}(x) = f^{k+j} \circ G^{-1}(x)$. We may verify this by remembering that G^{-1} is one-to-one on U_1 and for $l = 1,\ldots,k$ the values of x^l are $f^l(u)$, $u \in G^{-1}(U_1)$, so $f^l \circ G^{-1}(x) = f^l(u) = x^l$. So far we have used only the fact that the rank of DF at a (hence in a neighborhood of a) is *at least* k. We have not used the fact that it is identically k on A_0, but we need this in the next step which requires that the rank be *at most* k. We compute $D(F \circ G^{-1})$ from the formula above for $F \circ G^{-1}$, giving

$$D(F \circ G^{-1})(x) = \left(\begin{array}{c|ccc} I_k & & 0 & \\ \hline & \dfrac{\partial \bar{f}^{k+1}}{\partial x^{k+1}} & \cdots & \dfrac{\partial \bar{f}^{k+1}}{\partial x^n} \\ * & \vdots & & \vdots \\ & \dfrac{\partial \bar{f}^m}{\partial x^{k+1}} & \cdots & \dfrac{\partial \bar{f}^m}{\partial x^n} \end{array}\right).$$

This is valid on U_1, where $F \circ G^{-1}$ is defined. On the other hand, DG^{-1} is nonsingular on U_1 and $G^{-1}(U_1) = A_1 \subset A_0$. Therefore,

$$\operatorname{rank} D(F \circ G^{-1}) = \operatorname{rank}(DF \circ DG^{-1}) \equiv k$$

on U_1, which implies that all terms in the lower right-hand block of the matrix are zero on U_1, that is, the functions $\bar{f}^{i+1}, \ldots, \bar{f}^m$ depend on $x^1, \ldots, x^k$ only.

Now we define a function T from a neighborhood V_1 of 0 in $\boldsymbol{R}^m$ into $B_0 \subset \boldsymbol{R}^m$ by the formula

$$T(y^1, \ldots, y^k, y^{k+1}, \ldots, y^m) = (y^1, \ldots, y^k, y^{k+1} + \bar{f}^{k+1}(y^1, \ldots, y^k), \\ \ldots, y^m + \bar{f}^m(y^1, \ldots, y^k)).$$

The domain V_1 is subject to two restrictions, first it must be small enough so that for $y = (y^1, \ldots, y^m) \in V$, the functions $\bar{f}^{k+j}(y^1, \ldots, y^k)$ are defined and second, small enough so that $T(V_1) \subset B_0$. It is clear that $T(0) = 0$. If we compute DT, we see that it is nonsingular everywhere on V_1 since it takes the form

$$DT(y) = \left(\begin{array}{c|c} I_k & 0 \\ \hline * & I_{m-k} \end{array}\right).$$

Therefore T is a C^r diffeomorphism of a neighborhood V of 0 in V_1 onto an open set $B \subset \boldsymbol{R}^m$; the origin of $\boldsymbol{R}^m$ is in B and $B \subset B_1$. Choose a neighborhood $U \subset U_1$ of the origin in $\boldsymbol{R}^n$ such that $F \circ G^{-1}(U) \subset B$; let $A = G^{-1}(U)$ and let $H = T^{-1}$. Then

$$U \xrightarrow{G^{-1}} A \xrightarrow{F} B \xrightarrow{H} V$$

are C^r maps of these open sets and G^{-1}, H are C^r diffeomorphisms onto A and V, respectively. Finally we see that

$$H \circ F \circ G^{-1}(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0) \in \boldsymbol{R}^m$$

since

$$F \circ G^{-1}(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) = (x^1, \ldots, x^k, \bar{f}^{k+1}(x^1, \ldots, x^k), \\ \ldots, \bar{f}^m(x^1, \ldots, x^k)).$$

On the other hand, according to its definition above, T must take the value $(x^1, \ldots, x^k, \bar{f}^{k+1}(x^1, \ldots, x^k), \ldots, \bar{f}^m(x^1, \ldots, x^k))$ if we set $y^i = x^i$ for $i = 1, \ldots, k$ and $y^i = 0$ for $i = k + 1, \ldots, m$. Because T is one-to-one, it follows that T^{-1} takes $(x^1, \ldots, x^k, \bar{f}^{k+1}(x^1, \ldots, x^m), \ldots, \bar{f}^m(x^1, \ldots, x^m))$ to

(7.2) Corollary *We may choose the neighborhoods U and V in either of the following ways:* (i) $U = B_\varepsilon^n(0)$ *and* $V = B_\varepsilon^m(0)$ *or* (ii) $U = C_\varepsilon^n(0)$ *and* $V = C_\varepsilon^m(0)$ *with the same* $\varepsilon > 0$ *for both* U *and* V. *Then if* π *denotes the projection of* $\boldsymbol{R}^m = \boldsymbol{R}^k \times \boldsymbol{R}^{m-k}$ *to* $\boldsymbol{R}^k$ *and* $i: \boldsymbol{R}^k \to \boldsymbol{R}^k \times \boldsymbol{R}^{n-k}$ *is the injection to the subset* $\boldsymbol{R}^k \times \{0\}$, *we have* $\pi \circ H \circ F \circ G^{-1} \circ i$ *is the identity on* $B_\varepsilon^k(0)$ *in case* (i) *or on* $C_\varepsilon^k(0)$ *in case* (ii).

Exercises

1. Let $H^n \subset \boldsymbol{R}^n$ be defined by $H^n = \{x \mid x^n \geq 0\}$ and let $\partial H^n = \{x \mid x^n = 0\}$, x^n being the last coordinate of $x = (x^1, \ldots, x^n)$. We see that $\partial H^n = \boldsymbol{R}^{n-1}$. Suppose that U, V are (relatively) open sets of H^n and $F: U \to V$, $G: V \to U$ are C^r maps and are inverses of one another. Show that if $U' = U \cap \partial H^n$, $V' = V \cap \partial H^n$, we must have that $F \mid U'$ and $G \mid V'$ are one-to-one onto, inverses of each other, and C^r maps.
2. Prove Corollary 7.2.

We say that m functions $f^1(x), \ldots, f^m(x)$ of class C^1 defined on an open subset $U \subset \boldsymbol{R}^n$ are *dependent* if there exists a C^1 function $F(y^1, \ldots, y^m)$ on $\boldsymbol{R}^m$ which does not vanish on any open subset but such that $F(f^1(x), \ldots, f^m(x)) \equiv 0$ on U.

3. Show that if $f^1(x), \ldots, f^m(x)$ are dependent, then the rank of $\partial(f^1, \ldots, f^m)/\partial(x^1, \ldots, x^n)$ is less than m. Also as a partial converse show that if the rank is less than m and *constant* on U, then the functions are dependent locally.

Notes

The implicit function theorem, which is proved in many advanced calculus texts, is essentially equivalent to the inverse function theorem proved in the last section. For a proof of this latter theorem, which is not based on the contracting mapping theorem, see Spivak [1]. We have used the form of proof above, since this same principle may be applied to give a proof of the existence of solutions of systems of ordinary differential equations. All of these theorems are treated in a unified and very elegant way by Dieudonné [1], although there is a disadvantage for many readers in the fact that Banach space, rather than $\boldsymbol{R}^n$ is used throughout. S. Lang [1] also presents a very good treatment along the same lines. A discussion of the contracting mapping theorem and a sketch of its use in proving the existence theorem for differential equations may be found in the work of Kaplansky [1].

Although many of the theorems found here are valid for C^r functions and mappings and for C^ω also, the latter is too restrictive for most of our needs (theorems like those of Section 5 do not hold), and C^r is not strong enough to make Lemma 4.3 hold, which means that the characterization of $T_a(\boldsymbol{R}^n)$ given in Section 4 would have to be abandoned. For this reason, and since it is very convenient to know that we do not lose differentiability as a result of taking derivatives, i.e., the derivatives of a C^∞ function are also C^∞, C^∞ is the preferred differentiability class in much of differentiable manifold theory. We shall therefore consider functions and mappings of class C^∞ almost exclusively hereafter.

III DIFFERENTIABLE MANIFOLDS AND SUBMANIFOLDS

In the first section we give a precise definition of a C^∞ manifold of dimension n: a topological manifold together with a covering by compatible coordinate neighborhoods, that is, a covering such that a change of local coordinates is given by C^∞ mappings in $\boldsymbol{R}^n$. Several examples are worked out in detail, the most complicated being the Grassmann manifold of k-planes through the origin of $\boldsymbol{R}^n$.

In Sections 3 and 4 both C^∞ functions and mappings of manifolds are defined in terms of local coordinates, as is the rank of a mapping, that is, the rank of the Jacobian (in local coordinates). This enables us to consider certain examples of mappings of maximum rank (immersions and imbeddings) and leads to the definition of submanifold and regular submanifold (Section 5). We require the latter to be subspaces and to be defined locally by the vanishing of some of the coordinates of suitable local coordinates in the ambient space.

In Section 6 the concept of Lie group is defined. These are groups which are C^∞ manifolds such that the group operations are C^∞ mappings. It is shown that some standard matrix groups are Lie groups, for example, the group of orthogonal $n \times n$ matrices. In the following section the action of a Lie group on a manifold is defined and discussed. The group of rigid motions of $\boldsymbol{E}^n$ is an example.

The chapter concludes with a discussion of the special case of the properly discontinuous action of a discrete Lie group on a manifold and a brief introduction to covering manifolds.

1 The Definition of a Differentiable Manifold

As a preliminary to the definition of a differentiable manifold, we recall the definition of a topological manifold M of dimension n; it is a Hausdorff space with a countable basis of open sets and with the further property that each point has a neighborhood homeomorphic to an *open* subset of $\boldsymbol{R}^n$. Each pair U, φ, where U is an open set of M and φ is a homeomorphism of U to an open subset of $\boldsymbol{R}^n$, is called a *coordinate neighborhood*: to $q \in U$ we assign the n *coordinates* $x^1(q), \ldots, x^n(q)$ of its image $\varphi(q)$ in $\boldsymbol{R}^n$—each $x^i(q)$ is a real-valued function on U, the ith *coordinate function*. If q lies also in a second coordinate neighborhood V, ψ, then it has coordinates $y^1(q), \ldots, y^n(q)$ in this neighborhood. Since φ and ψ are homeomorphisms, this defines a homeomorphism

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V),$$

the domain and range being the two open subsets of $\boldsymbol{R}^n$ which correspond to the points of $U \cap V$ by the two coordinate maps φ, ψ, respectively. In coordinates, $\psi \circ \varphi^{-1}$ is given by continuous functions

$$y^i = h^i(x^1, \ldots, x^n), \qquad i = 1, \ldots, n,$$

giving the y-coordinates of each $q \in U \cap V$ in terms of its x-coordinates. Similarly $\varphi \circ \psi^{-1}$ gives the inverse mapping which expresses the x-coordinates as functions of the y-coordinates

$$x^i = g^i(y^1, \ldots, y^n), \qquad i = 1, \ldots, n.$$

The fact that $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ are homeomorphisms and are inverse to each other is equivalent to the continuity of $h^i(x)$ and $g^j(y)$, $i, j = 1, \ldots, n$ together with the identities

$$h^i(g^1(y), \ldots, g^n(y)) \equiv y^i, \qquad i = 1, \ldots, n,$$

and

$$g^j(h^1(x), \ldots, h^n(x)) \equiv x^j, \qquad j = 1, \ldots, n.$$

Thus every point of a topological manifold M lies in a very large collection of coordinate neighborhoods, but whenever two neighborhoods overlap we have the formulas just given for change of coordinates. The basic idea that leads to differentiable manifolds is to try to select a family or subcollection of neighborhoods so that the change of coordinates is always given by differentiable functions.

(1.1) Definition We shall say that U, φ and V, ψ are C^∞-*compatible* if $U \cap V$ nonempty implies that the functions $h^i(x)$ and $g^j(y)$ giving the change of coordinates are C^∞; this is equivalent to requiring $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ to be *diffeomorphisms* of the open subsets $\varphi(U \cap V)$ and $\psi(U \cap V)$ of $\boldsymbol{R}^n$. (See Fig. III.1.)

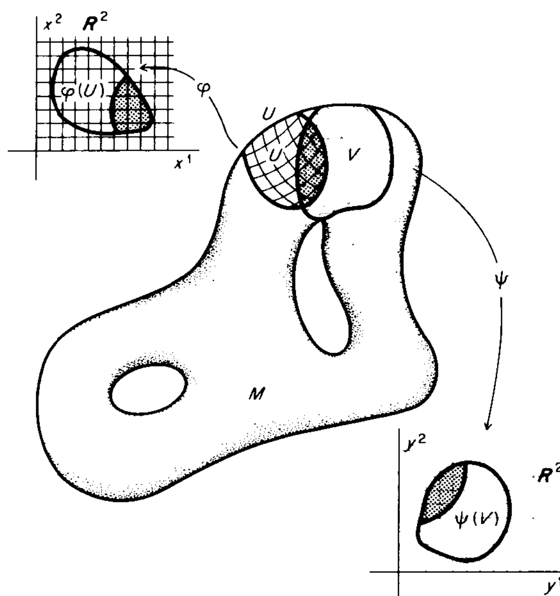


Figure III.1

(1.2) Definition A *differentiable or* C^∞ (*or smooth*) *structure* on a topological manifold M is a family $\mathscr{U} = \{U_\alpha, \varphi_\alpha\}$ of coordinate neighborhoods such that:

(1) the U_α cover M,

(2) for any α, β the neighborhoods U_α, φ_α and U_β, φ_β are C^∞-compatible,

(3) any coordinate neighborhood V, ψ compatible with every $U_\alpha, \varphi_\alpha \in \mathscr{U}$ is itself in $\mathscr{U}$.

A C^∞ *manifold* is a topological manifold together with a C^∞-differentiable structure.

It is, of course, conceivable that for some topological manifold no such family of compatible coordinate neighborhoods can be singled out. It is also conceivable that, on the contrary, families can be chosen in a multiplicity of inequivalent ways so that two inequivalent C^∞ manifolds have the same underlying topological manifold. These are basic but very difficult questions, and in fact, are matters of recent research (see the notes at the end of this chapter). What is important from our point of view is that we will be able to find an abundance of topological manifolds with at least one differentiable structure, thus an abundance of C^∞ manifolds; so we may ignore these more difficult questions in what we do here.

Since there is no danger of confusion, we will often say simply "manifold" for C^∞ manifold; we may also sometimes say differentiable or smooth manifold. Moreover, "coordinate neighborhood" will hereafter refer exclusively to the *coordinate neighborhoods belonging to the differentiable structure.* Should we have occasion to consider a manifold without differentiable structure, we will say *topological* manifold and *topological* coordinate neighborhood.

By requiring only that the change of coordinates be given by C^r functions for $r < \infty$, we could define C^r-compatible coordinate neighborhoods and C^r manifolds, C^0 denoting a topological manifold. One can also require that the change of coordinates be C^ω, that is, real-analytic. We shall restrict ourselves almost exclusively to the C^∞ case.

Before proceeding we will prove the following proposition, which will make it easier to give examples of differentiable manifolds; it shows that (1) and (2) of Definition 1.2 are the essential properties defining a C^∞ structure. Thus in examples we need only check the compatibility of a covering by neighborhoods.

(1.3) Theorem *Let M be a Hausdorff space with a countable basis of open sets. If $\{V_\beta, \psi_\beta\}$ is a covering of M by C^∞-compatible coordinate neighborhoods, then there is a unique C^∞ structure on M containing these coordinate neighborhoods.*

Proof We shall define the differentiable structure to be the collection $\mathscr{U}$ of all topological coordinate neighborhoods U, φ which are C^∞-compatible with each and every one of those of the given collection $\{V_\beta, \psi_\beta\}$. This new collection naturally includes the V_β, ψ_β and so property (1) of Definition 1.2 is automatically satisfied. As to property (2), suppose U, φ and U', φ', $U \cap U' \neq \varnothing$, are in the collection we have defined. Then are they C^∞-compatible? Since they are (topological) coordinate neighborhoods, the functions $\varphi' \circ \varphi^{-1}$ and $\varphi \circ \varphi'^{-1}$ giving the change of coordinates are well-defined homeomorphisms on open subsets of $\boldsymbol{R}^n$, and we need only be sure that they are C^∞. Let $x = \varphi(p)$ be an arbitrary point of $\varphi(U \cap U')$. Then $p \in V_\beta$ for one of the coordinate neighborhoods V_β, ψ_β. Therefore $W = V_\beta \cap U \cap U'$ is an open set containing p, and $\varphi(W)$ is an open set containing x. We have $\varphi' \circ \varphi^{-1} = \varphi' \circ \psi_\beta^{-1} \circ \psi_\beta \circ \varphi^{-1}$ on $\varphi(W)$, but $\varphi' \circ \psi^{-1}$ and $\psi_\beta \circ \varphi^{-1}$ are C^∞ since U, φ and U', φ' are both C^∞-compatible with V_β, ψ_β. It follows that their composition $\varphi' \circ \varphi^{-1}$ is C^∞ on $\varphi(W)$; and since it is C^∞ on a neighborhood of any point of its domain, it is C^∞. This proves everything except property (3), which is automatic: Any U, φ that is compatible with all of the coordinate neighborhoods in our collection certainly has this property with respect to the subcollection $\{V_\beta, \psi_\beta\}$, and is thus in the differentiable structure.

(1.4) Remark It is important to note that a coordinate neighborhood U, φ depends on *both* the neighborhood U and the map φ of U to $\boldsymbol{R}^n$. If we change either, then we have a different coordinate neighborhood. For example, if $V \subset U$ is an open subset, then $V, \varphi \mid V$ is a new coordinate neighborhood—technically at least—although the coordinates of $p \in V$ are the same as its coordinates in the original neighborhood. If $p \in U$, we may always choose V so that $\varphi(V)$ is an open ball $B_\varepsilon^n(a)$, or cube $C_\varepsilon^n(a)$, in $\boldsymbol{R}^n$ with $\varphi(p) = a$ as center. We may further alter φ by composing it with a translation $\theta\colon \boldsymbol{R}^n \to \boldsymbol{R}^n$ taking $\varphi(p) = a$ to $(0, 0, \ldots, 0)$. Then V, ψ with $\psi = \theta \circ \varphi \mid V$ is a new coordinate system, $p \in V \subset U$, $\psi(p) = (0, 0, \ldots, 0)$, and $\psi(V) = B_\varepsilon^n(0)$ or $C_\varepsilon^n(0)$.

Using the theorem just proved, we give some preliminary examples of manifolds.

(1.5) Example (*The Euclidean plane*) (See comments in Section I.2.) Once a unit of length is chosen, the Euclidean plane $\boldsymbol{E}^2$ becomes a metric space. It is Hausdorff and has a countable basis of open sets; the choice of an origin and mutually perpendicular coordinate axes establishes a homeomorphism (even an isometry) $\psi : \boldsymbol{E}^2 \to \boldsymbol{R}^2$. Thus we cover $\boldsymbol{E}^2$ with a single coordinate neighborhood V, ψ with $V = \boldsymbol{E}^2$ and $\psi(V) = \boldsymbol{R}^2$. It follows not only that $\boldsymbol{E}^2$ is a topological manifold, but by Theorem 1.3, V, ψ determines a differentiable structure, so $\boldsymbol{E}^2$ is a C^∞ manifold.

There are many other coordinate neighborhoods on $\boldsymbol{E}^2$ which are C^∞-compatible with V, ψ, that is, which belong to the differentiable structure determined by V, ψ. For example, we may choose another rectangular Cartesian coordinate system V', ψ'. Then it is shown in analytic geometry that the change of coordinates is given by linear, hence C^∞ (even analytic!) functions

$$y^1 = x^1 \cos\theta - x^2 \sin\theta + h, \qquad y^2 = x^1 \sin\theta + x^2 \cos\theta + k.$$

Note that $V = V'$, but the coordinate neighborhoods are *not* the same since $\psi' \neq \psi$, that is, the coordinates of each point are different for the two mappings.

It is also possible to choose as U the plane minus a ray extending from a point 0. Using the angle $\theta(q)$ measured from this ray to $\overline{0q}$ and the distance $r(q)$ of q from 0 as coordinate functions on U we define a homeomorphism $\varphi(q) = (r(q), \theta(q))$ from U to the open set $\{(r, \theta) \mid r > 0, 0 < \theta < 2\pi\}$ in $\boldsymbol{R}^2$. The equations for change of coordinates to those above, assuming that 0 is the origin and that the ray is the positive x-axis, are

$$x^1 = r\cos\theta, \qquad x^2 = r\sin\theta,$$

which again are analytic, thus C^∞. If the origin and axes are not chosen in this special way, then we must compose this mapping on $\boldsymbol{R}^2$ with a rotation

and translation of the type above to obtain the functions giving the change of coordinates. The various coordinate neighborhoods just enumerated being C^∞-compatible with our original V, ψ are in the differentiable structure on $\boldsymbol{E}^2$ determined by it.

In the same manner Euclidean space of arbitrary dimension n gives an example of a C^∞ manifold, covered by a single coordinate system. Again, this may be done in a variety of ways. As we have noted it is customary to identify $\boldsymbol{E}^n$ and $\boldsymbol{R}^n$ since the former is difficult to axiomatize; this is equivalent to choosing a *fixed* rectangular Cartesian coordinate system covering all of $\boldsymbol{E}^n$. Many examples will make it abundantly clear that manifolds in general can *not* be covered by a single coordinate system nor are there preferred coordinates. Thus it is often better in thinking of Euclidean space as a *manifold* to visualize the model $\boldsymbol{E}^n$ of classical geometry—without coordinates—rather than $\boldsymbol{R}^n$, Euclidean space with coordinates. (However, we will later follow common practice and identify $\boldsymbol{E}^n$ and $\boldsymbol{R}^n$.)

A finite-dimensional vector space V over $\boldsymbol{R}$ can be identified with $\boldsymbol{R}^n$, $n = \dim V$, once a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is chosen: $\mathbf{v} = x^1\mathbf{e}_1 + \cdots + x^n\mathbf{e}_n$ is identified with $(x^1, \ldots, x^n)$ in $\boldsymbol{R}^n$; similarly, the $m \times n$ matrices (a_{ij}) with $\boldsymbol{R}^{mn}$ with the matrix $A = (a_{ij})$ corresponding to

$$(a_{11}, \ldots, a_{1n}; \ldots; a_{m1}, \ldots, a_{mn}).$$

Using these identification mappings we may define a natural topology and C^∞ structure on V and on the set $\mathscr{M}_{mn}(\boldsymbol{R})$ of $m \times n$ matrices over $\boldsymbol{R}$. We suppose them to be homeomorphic to Cartesian or Euclidean space of dimension n in the case of V, and mn in the case of $\mathscr{M}_{mn}(\boldsymbol{R})$ and covered by a single coordinate neighborhood, the identification map above being the coordinate map.

(1.6) Example (*Open submanifolds*) An open subset U of a C^∞ manifold M is itself a C^∞ manifold with differentiable structure consisting of the coordinate neighborhoods V', ψ' obtained by restriction of ψ, on those coordinate neighbohoods V, ψ which intersect U, to the open set $V' = V \cap U$, that is, $\psi' = \psi | V \cap U$. This gives a covering of U by topological coordinate neighborhoods which are C^∞-compatible, and hence defines a C^∞ structure on U. which is said then to be an *open submanifold* of M.

A particular case of some interest is the following. We consider the subset $U = Gl(n, \boldsymbol{R})$ of $M = \mathscr{M}_n(\boldsymbol{R})$, $n \times n$ matrices over $\boldsymbol{R}$, which consists of all nonsingular $n \times n$ matrices. Since an $n \times n$ matrix A is nonsingular if and only if its determinant $\det A$ is not zero, we have

$$U = \{A \in \mathscr{M}_n(\boldsymbol{R}) \mid \det A \neq 0\},$$

which is the usual definition of the group $Gl(n, \boldsymbol{R})$. Since $\det A$ is a polynomial function of its entries a_{ij}, it is a continuous function of its entries and of

A in the topology of identification with R^{n^2}. Thus $U = Gl(n, R)$ is an open set—the complement of the closed set of those A such that $\det A = 0$, and we see that $Gl(n, R)$ is an open submanifold of $\mathcal{M}_n(R)$.

(1.7) Theorem *Let M and N be C^∞ manifolds of dimensions m and n. Then $M \times N$ is a C^∞ manifold of dimension $m + n$ with C^∞ structure determined by coordinate neighborhoods of the form $\{U \times V, \varphi \times \psi\}$, where U, φ and V, ψ are coordinate neighborhoods on M and N, respectively, and $(\varphi \times \psi)(p, q) = (\varphi(p), \psi(q))$ in $R^{m+n} = R^m \times R^n$.*

The proof is left as an exercise. An important example is the torus $T^2 = S^1 \times S^1$, the product of two circles (see Fig. III.2). More generally, $T^n = S^1 \times \cdots \times S^1$, the n-fold product of circles is a C^∞ manifold obtained as a Cartesian product.

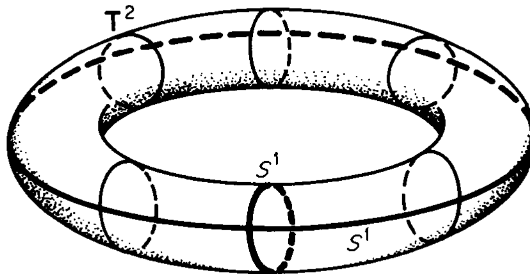


Figure III.2

(1.8) Example (*The Sphere*) We give a fairly detailed proof, using Theorem 1.3, that the unit 2-sphere $S^2 = \{x \in R^3 \mid \|x\| = 1\}$ is a C^∞ manifold (see Fig. III.3). The idea used is an elaboration of that discussed in Section I.3. It extends in an obvious way to S^{n-1}, the unit $n - 1$ sphere in R^n. A somewhat simpler method, using stereographic projection, is left to the exercises; it also extends to S^{n-1}.

We take S^2 with its topology as a subspace of R^3, that is, U is open in S^2 if $U = \tilde{U} \cap S^2$ for some open set $\tilde{U} \subset R^3$. This implies that S^2 is Hausdorff with a countable basis; we shall show that it is locally Euclidean. For $i = 1$, 2, or 3, let $\tilde{U}_i^+ = \{(x^1, x^2, x^3) \mid x^i > 0\}$ and $\tilde{U}_i^- = \{(x^1, x^2, x^3) \mid x^i < 0\}$; these $\tilde{U}_i^\pm$ are two open sets into which the coordinate hyperplane $x_i = 0$ divides R^3. The relatively open sets $U_i^\pm = \tilde{U}_i^\pm \cap S^2$, $i = 1, 2, 3$, cover S^2. We define $\varphi_i^\pm : U_i^\pm \to R^2$ by projection:

$$\varphi_1^\pm(x^1, x^2, x^3) = (x^2, x^3),$$
$$\varphi_2^\pm(x^1, x^2, x^3) = (x^1, x^3),$$
$$\varphi_3^\pm(x^1, x^2, x^3) = (x^1, x^2).$$

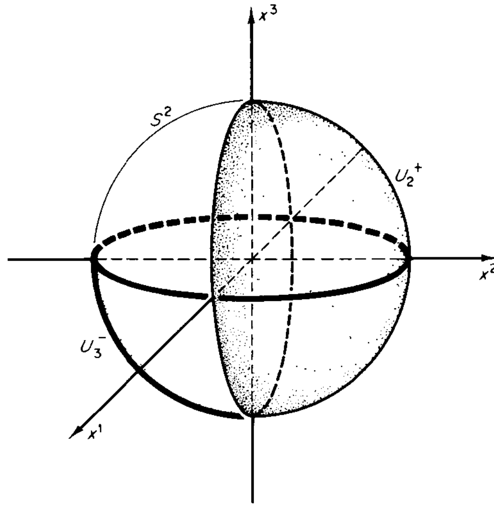


Figure III.3

These are homeomorphisms to the open set $W = \{x \in \boldsymbol{R}^2 \mid \|x\| < 1\}$ as is easily checked; thus S^2 is locally Euclidean and a topological manifold. However, the formulas for the change of coordinates are C^∞, and thus these coordinate neighborhoods are C^∞-compatible. For example, $\varphi_1^+ \circ (\varphi_2^-)^{-1}$ is given on $U_1^+ \cap U_2^-$ by composing $(\varphi_2^-)^{-1}$ and φ_1^+

$$(x^1, x^3) \xrightarrow{(\varphi_2^-)^{-1}} (x^1, -(1 - (x^1)^2 - (x^3)^2)^{1/2}, x^3)$$

$$(x^1, -(1 - (x^1)^2 - (x^3)^2)^{1/2}, x^3) \xrightarrow{\varphi_1^+} (-(1 - (x^1)^2 - (x^3)^2)^{1/2}, x^3).$$

Then, by change of notation, using (u^1, u^2) as U_2^--coordinates and (v^1, v^2) as U_1^+-coordinates instead of (x^1, x^3) and (x^2, x^3), we have

$$v^1 = -(1 - (u^1)^2 - (u^2)^2)^{1/2}, \qquad v^2 = u^2.$$

The v^i are C^∞ functions of the u^i since the square root term is never zero on $\{(u^1, u^2) | (u^1)^2 + (u^2)^2 < 1\}$. Similarly, $\varphi_2^- \circ (\varphi_1^+)^{-1}$ is C^∞ on $\{(v^1, v^2) | (v^1)^2 + (v^2)^2 < 1\}$. Thus the coordinate neighborhoods U_1^+, φ_1^+ and U_2^-, φ_2^- are C^∞-compatible. This covering of S^2 by eight coordinate neighborhoods determines a C^∞ structure.

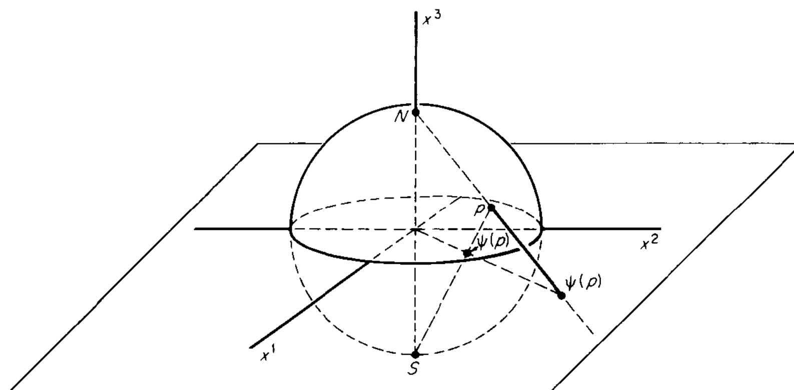
Thus S^2 is an example of a manifold which is a subset of another manifold, namely $\boldsymbol{R}^3$. Very many examples will be of this type as will be seen later; they are called *submanifolds* (to be defined). A two-dimensional submanifold

of E^3 or R^3 is often called a *surface* and a one-dimensional submanifold a *curve*; planes and spheres, circles and lines are the simplest examples. Classical differential geometry dealt extensively with these two cases.

Coordinate neighborhoods are often called *charts* and a collection of compatible charts which cover a manifold M an *atlas*, a terminology suggested by charts or maps of the globe. An atlas thus determines a differentiable structure on M. A *complete atlas* is maximal, i.e., satisfies (3) of Definition 1.2.

Exercises

1. Prove Theorem 1.7.
2. Using stereographic projection from the north pole $N(0, 0, +1)$ of all of the standard unit sphere in R^3 except $(0, 0, -1)$ determine a coordinate neighborhood U_N, φ_N. In the same way determine by projection from the south pole $S(0, 0, -1)$ a neighborhood U_S, φ_S (see the accompanying figure). Show that these two neighborhoods determine a C^∞ structure on S^2. Generalize to S^{n-1}.



3. Let $\mathscr{M}_{mn}(\boldsymbol{R})$ be the space of all real $m \times n$ matrices and $\mathscr{M}_{mn}^k(\boldsymbol{R})$ be the subset of all those $m \times n$ matrices whose rank is $\geq k$. Show that $\mathscr{M}_{mn}^k(\boldsymbol{R})$ is an open subset of $\mathscr{M}_{mn}(\boldsymbol{R})$ and thus a manifold.
4. We say that two (C^∞) atlases on M are *equivalent* if their union is an atlas. Show that this is an equivalence relation and that each equivalence class contains a unique complete atlas.
5. Let X be a set and $\{U_\alpha, \varphi_\alpha\}$ a covering by subsets U_α each with a bijective map $\varphi_\alpha: U_\alpha \to V_\alpha$, an open subset of $\boldsymbol{R}_n$. If these charts are all C^∞ compatible, i.e., all $\varphi_\beta \circ \varphi_\alpha^{-1}$ are C^∞, does this data define a unique manifold topology—with φ_α^{-1}-images of open subsets of V_α for all α as basis of open sets—and a corresponding C^∞ structure on X?

2 Further Examples

In this section we shall discuss two related examples of "abstract" manifolds, that is, manifolds which are not defined as submanifolds of Euclidean space. The first example is the space of classical (real) projective geometry, the second—technically more difficult to define than any example we have met thus far—is the Grassman manifold, consisting of all k-planes through the origin in $\boldsymbol{R}^n$, $n > k$; it will be treated more fully in Section IV.9. Both examples—in fact the latter includes the former—arise from equivalence relations defined on simpler manifolds, the underlying space of the new manifold being the set of equivalence classes with a suitable topology.

Let X be a topological space and $\sim$ an equivalence relation on X. Denote by $[x] = \{y \in X \mid y \sim x\}$ the equivalence class of x, and for a subset $A \subset X$, denote by $[A]$ the set $\bigcup_{a \in A} [a]$, that is, all x equivalent to some element of A. We let $X/\sim$ stand for the set of equivalence classes and denote by $\pi : X \to X/\sim$ the natural mapping (projection) taking each $x \in X$ to its equivalence class, $\pi(x) = [x]$. With these notations we define the standard *quotient topology on* $X/\sim$ as follows: $U \subset X/\sim$ is an open subset if $\pi^{-1}(U)$ is open; the projection π is then continuous.

(2.1) Definition With the above notation and topology we shall call $X/\sim$ the *quotient space* of X relative to the relation $\sim$.

As a simple example let $X = \boldsymbol{R}$ the real numbers and let $\boldsymbol{Z}$ be the integers. We define $x \sim y$ if $x - y \in \boldsymbol{Z}$ and denote by $\boldsymbol{R}/\sim$ the quotient space. We shall leave as an exercise the proof that this quotient space may be naturally identified with $S^1 = \{z \in C \mid |z| = 1\}$, the unit circle in the complex plane, and that $\pi : \boldsymbol{R} \to \boldsymbol{R}/\sim$ is then identified with the map $\pi(t) = \exp(2\pi t\sqrt{-1})$. Note that $X/\sim$ is a space of cosets of a group relative to a subgroup; this situation occurs frequently.

(2.2) Definition An equivalence relation $\sim$ on a space X is called *open* if whenever a subset $A \subset X$ is open, then $[A]$ is also open.

Our examples will usually be open equivalence relations. The following lemma will show why.

(2.3) Lemma *An equivalence relation $\sim$ on X is open if and only if π is an open mapping. When $\sim$ is open and X has a countable basis of open sets, then $X/\sim$ has a countable basis also.*

Proof Let $A \subset X$ be an open subset. Since $[A] = \pi^{-1}(\pi(A))$, we see by definition of the quotient topology on $X/\sim$ that $[A]$ is open if π is open and

conversely $[A]$ open implies $\pi(A)$ is open. Now suppose $\sim$ is open and X has a countable basis $\{U_i\}$ of open sets. If W is an open subset of $X/\sim$, then $\pi^{-1}(W) = \bigcup_{j \in J} U_j$ for some subfamily of $\{U_i\}$ and $W = \pi(\pi^{-1}(W)) = \bigcup_{j \in J} \pi(U_j)$. It follows that $\{\pi(U_i)\}$ is a basis of open sets for $X/\sim$. ∎

This lemma is clearly useful in determining those equivalence relations on a manifold M whose quotient space is again a manifold, for a manifold must be a Hausdorff space with a countable basis of open sets. Unfortunately, there is no simple condition which will assure that the quotient space is Hausdorff. In fact, as Exercise 2 shows, a quotient space $X/\sim$ may be locally Euclidean with a countable basis of open sets and still fail to be Hausdorff. Nevertheless we obtain important examples by this method, sometimes with the assistance of the following lemma.

(2.4) Lemma *Let $\sim$ be an open equivalence relation on a topological space X. Then $R = \{(x, y) \mid x \sim y\}$ is a closed subset of the space $X \times X$ if and only if the quotient space $X/\sim$ is Hausdorff.*

Proof Suppose $X/\sim$ is Hausdorff and suppose $(x, y) \notin R$, that is, $x \not\sim y$. Then there are disjoint neighborhoods U of $\pi(x)$ and V of $\pi(y)$. We denote by $\tilde{U}$ and $\tilde{V}$ the open sets $\pi^{-1}(U)$ and $\pi^{-1}(V)$, which contain x and y, respectively. If the open set $\tilde{U} \times \tilde{V}$ containing (x, y) intersects R, then it must contain a point (x', y') for which $x' \sim y'$, so that $\pi(x') = \pi(y')$ contrary to the assumption that $U \cap V = \varnothing$. This contradiction shows that $\tilde{U} \times \tilde{V}$ does not intersect R and that R is closed.

Conversely, suppose that R is closed, then given any distinct pair of points $\pi(x)$, $\pi(y)$ in $X/\sim$, there is an open set of the form $\tilde{U} \times \tilde{V}$ containing (x, y) and having no point in R. It follows that $U = \pi(\tilde{U})$ and $V = \pi(\tilde{V})$ are disjoint. Lemma 2.3 and the hypothesis imply that U and V are open. Thus $X/\sim$ is Hausdorff. ∎

(2.5) Example (*Real projective space $P^n(\boldsymbol{R})$*) We let $X = \boldsymbol{R}^{n+1} - \{0\}$, all $(n+1)$-tuples of real numbers $x = (x^1, \ldots, x^{n+1})$ except $0 = (0, \ldots, 0)$, and define $x \sim y$ if there is a real number $t \neq 0$ such that $y = tx$, that is,

$$(y_1, \ldots, y_{n+1}) = (tx_1, \ldots, tx_{n+1}).$$

The equivalence classes $[x]$ may be visualized as lines through the origin (Fig.III.4). We denote the quotient space by $P^n(\boldsymbol{R})$; it is called *real projective space.* We prove:

$P^n(\boldsymbol{R})$ is a differentiable manifold of dimension n.

To do so we first note that $\pi : X \to P^n(\boldsymbol{R})$ is an open mapping. If $t \neq 0$ is a real number, let $\varphi_t : X \to X$ be the mapping defined by $\varphi_t(x) = tx$. It is clearly a homeomorphism with $\varphi_t^{-1} = \varphi_{1/t}$. If $U \subset X$ is an open set, then

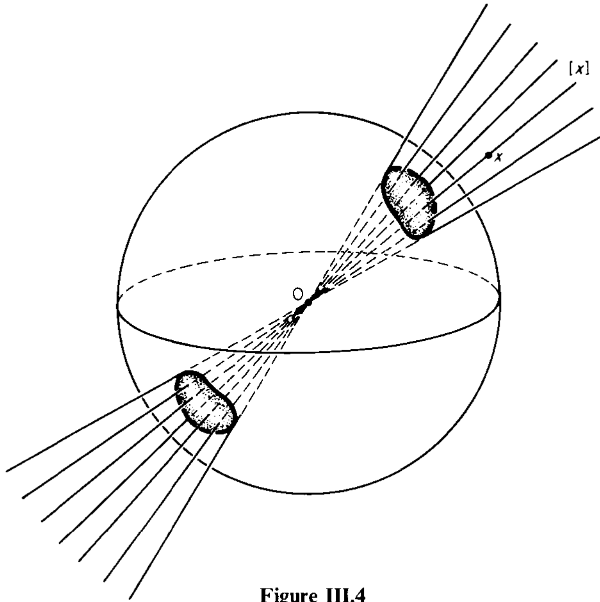


Figure III.4

$[U] = \bigcup \varphi_t(U)$, the union being over all real $t \neq 0$. Since each $\varphi_t(U)$ is open, $[U]$ is open and π is open by Lemma 2.3.

Next we apply Lemma 2.4 to prove that $P^n(\boldsymbol{R})$ is Hausdorff. On the open submanifold $X \times X \subset \boldsymbol{R}^{n+1} \times \boldsymbol{R}^{n+1}$ we define a real-valued function $f(x, y)$ by

$$f(x^1, \ldots x^{n+1}; y^1, \ldots, y^{n+1}) = \sum_{i \neq j} (x^i y^j - x^j y^i)^2.$$

Then $f(x, y)$ is continuous and vanishes if and only if $y = tx$ for some real number $t \neq 0$, that is, if and only if $x \sim y$. Thus

$$R = \{(x, y) \mid x \sim y\} = f^{-1}(0)$$

is a closed subset of $X \times X$ and $P^n(\boldsymbol{R})$ is Hausdorff.

We define $n + 1$ coordinate neighborhoods $U_i, \varphi_i, i = 1, \ldots, n + 1$, as follows: Let $\tilde{U}_i = \{x \in X \mid x^i \neq 0\}$ and $U_i = \pi(\tilde{U}_i)$. Then $\varphi_i \colon U_i \to \boldsymbol{R}^n$ is defined by choosing any $x = (x^1, \ldots, x^{n+1})$ representing $[x] \in U_i$ and putting

$$\varphi_i(x) = \left(\frac{x^1}{x^i}, \ldots, \frac{x^{i-1}}{x^i}, \frac{x^{i+1}}{x^i}, \ldots, \frac{x^{n+1}}{x^i}\right).$$

It is seen that if $x \sim y$, then $\varphi_i(x) = \varphi_i(y)$; moreover $\varphi_i(x) = \varphi_i(y)$ implies $x \sim y$. Thus $\varphi_i \colon U_i \to \boldsymbol{R}^n$ is properly defined, continuous, one-to-one, and even onto. For $z \in \boldsymbol{R}^n$, $\varphi_i^{-1}(z)$ is given by composing a C^∞ map of $\boldsymbol{R}^n$ to $\boldsymbol{R}^{n+1}$ with π, namely, $\varphi_i^{-1}(z^1, \ldots, z^n) = \pi(z^1, \ldots, z^{i-1}, +1, z^i, \ldots, z^n)$; therefore

φ_i^{-1} is continuous. Thus $P^n(\boldsymbol{R})$ is a (topological) manifold and is C^∞ if the coordinate neighborhoods are C^∞-compatible, that is, $\varphi_i \circ \varphi_j^{-1}$ is C^∞ (where defined) for $1 \leq i, j \leq n+1$. The verification is simple and explicit and is left to the reader. This completes the proof that $P^n(\boldsymbol{R})$ is a manifold.

(2.6) Example (*Grassman manifolds $G(k, n)$*) The Grassman manifold $G(k, n)$ is the set of all k-planes through the origin of $\boldsymbol{R}^n$- or k-dimensional subspaces of $V^n = \boldsymbol{R}^n$ (as a vector space)—endowed with a suitable topology and differentiable structure. We will realize $G(k, n)$ as a quotient space arising from an equivalence relation on the manifold $F(k, n)$ of k-frames in $\boldsymbol{R}^n$, where we define a *k-frame* in $\boldsymbol{R}^n$ to be a linearly independent set $\mathbf{x}$ of k elements of $\boldsymbol{R}^n$:

$$\begin{aligned} x_1 &= (x_1^1, \ldots, x_1^n), \\ &\vdots \\ x_k &= (x_k^1, \ldots, x_k^n). \end{aligned}$$

A k-frame in $\boldsymbol{R}^n$ may be identified with the $k \times n$ matrix, which we also denote by $\mathbf{x}$, whose rows are $x_1, \ldots, x_k$. We use the fact that the set $\mathcal{M}_{kn}(\boldsymbol{R})$ of all $k \times n$ real matrices is a differentiable manifold by virtue of its identification with $\boldsymbol{R}^{kn}$. The matrices which correspond to k-frames, that is, those of rank k, form an open subset and hence $F(k, n)$ is a differentiable manifold. This is because of the fact that "$\mathbf{x}$ is of rank k" means that the following two equivalent statements hold: (i) the row vectors form a linearly independent set and (ii) not all $k \times k$ minor determinants are zero simultaneously. Statement (ii) shows that the rank is less than k at the simultaneous zeros of a set of continuous functions on $\mathcal{M}_{kn}(\boldsymbol{R})$, that is, on a closed subset, so $F(k, n)$ is open.

Clearly each frame $\mathbf{x}$ determines a k-plane or point of $G(k, n)$, namely, the subspace spanned by $x_1, \ldots, x_k$, so that we have a natural map of $F(k, n)$ onto $G(k, n)$. Moreover $\mathbf{x} = (x_1, \ldots, x_k)$ and $\mathbf{y} = (y_1, \ldots, y_k)$ determine the same k-plane if and only if $y_i = \sum_{j=1}^k \alpha_{ij} x_j$, where $a = (\alpha_{ij})$ is a nonsingular $k \times k$ matrix, that is, if and only if $\mathbf{y} = a\mathbf{x}$, the product of the matrices a and $\mathbf{x}$. It is natural to define $\sim$ by

$$\mathbf{y} \sim \mathbf{x} \quad \text{if} \quad \mathbf{y} = a\mathbf{x}, \quad a \in Gl(k, \boldsymbol{R}).$$

We then identify $G(k, n)$ with $F(k, n)/\sim$, the set of equivalence classes, and the above mentioned natural map with π. We sketch a proof that $G(k, n)$ with the quotient space topology has the structure of a differentiable manifold of dimension $k(n-k)$. A different proof will be given in Section IV.9. Note that if $k = 1$, then $a \in G(1, \boldsymbol{R}) = \boldsymbol{R}^*$ and $G(k, n)$ becomes $P^{n-1}(\boldsymbol{R})$. The proof that π is an open mapping is analogous to Example 2.5 and is left to

the reader. The proof that $G(k, n)$ is Hausdorff is trickier, but is also left as an exercise.

It remains to describe a covering by coordinate neighborhoods with C^∞-compatible coordinate maps so that Theorem 1.3 may be applied to complete the proof. We shall use the $k \times k$ submatrices of $\mathbf{x} \in \mathscr{M}_{kn}(\boldsymbol{R})$ to accomplish this. Let $J = (j_1, \ldots, j_k)$ be an ordered subset of $(1, \ldots, n)$, for example, $J = (1, 2, \ldots, k)$ and J' be the complementary subset [for the example then, $J' = (k + 1, \ldots, n)$]. By $\mathbf{x}_J$ we denote the $k \times k$ submatrix $(x_i^{j_l})$, $1 \le i, l \le k$, of the $k \times n$ matrix $\mathbf{x}$, and by $\mathbf{x}_{J'}$ we denote the complementary $k \times (n - k)$ submatrix obtained by striking out the columns $j_1, \ldots, j_k$ of $\mathbf{x}$. Let $\tilde{U}_J$ be the open set in $F(k, n)$, consisting of matrices for which $\mathbf{x}_J$ is nonsingular and let $U_J = \pi(\tilde{U}_J)$ be the corresponding open set in $G(k, n)$. Each $\mathbf{y} \in \tilde{U}_J$ is equivalent to exactly one $k \times n$ matrix $\mathbf{x}$ in which the submatrix $\mathbf{x}_J$ is the $k \times k$ identity matrix; for example, if $J = (1, 2, \ldots, k)$, then $\mathbf{x}$ is of the form

$$\mathbf{x} = \begin{pmatrix} 1 & \cdots & 0 & x_{1,k+1} & \cdots & x_{1n} \\ 0 & \cdots & 0 & & & \\ \vdots & & \vdots & & & \vdots \\ 0 & \cdots & 1 & x_{k,k+1} & \cdots & x_{kn} \end{pmatrix}.$$

(In fact the $\mathbf{x}$ equivalent to a matrix $\mathbf{y}$ for which $\mathbf{y}_J$ is nonsingular is given by the matrix formula $\mathbf{x} = \mathbf{y}_J^{-1}\mathbf{y}$.)

We define $\varphi_J : U_J \to \mathscr{M}_{k(n-k)}(\boldsymbol{R})$, identified with $\boldsymbol{R}^{k(n-k)}$, by deleting the k columns corresponding to J in this representative $\mathbf{x}$ of $\mathbf{y}$, thus $\varphi_J([\mathbf{y}]) = \mathbf{x}_{J'}$ (the matrix comprising the last $n - k$ columns, in the example above). We leave it as an exercise to show that φ_J is properly defined and maps U_J onto $\boldsymbol{R}^{k(n-k)}$ homeomorphically and that the U_J, $\varphi_{J'}$ for all subsets J of k distinct elements of $(1, 2, \ldots, n)$, form a covering of $G(k, n)$ by C^∞-compatible coordinate neighborhoods; a verification of this for $G(2, 4)$, the 2-planes through the origin of $\boldsymbol{R}^4$, is sufficient to show how to proceed in general. As mentioned, a different proof will be given later.

Exercises

1. Prove the statements after Definition 2.1 concerning $\boldsymbol{R}/\sim$, S^1, and the mapping $\pi: \boldsymbol{R} \to \boldsymbol{R}/\sim$. Show that π is an open mapping.
2. Let $X = \boldsymbol{R} \times \{\pm 1\}$ with the Cartesian topology and C^∞ structure. Define an equivalence relation on X by $(x, i) \sim (y, j)$ if either (i) $x = y < 0$ or (ii) $x = y$ and $i = j$. Show that $X/\sim$ is locally Euclidean and has a countable basis of open sets but is not Hausdorff (cf. Exercise I.3.1).

3. Let $\alpha: S^n \to S^n$ be the map of the unit sphere in $\boldsymbol{R}^{n+1}$ taking each x to its antipodal point $\alpha(x) = -x$. Show that $x \sim y$ if $y = x$ or $y = \alpha(x)$ is an equivalence relation and that $S^n/\sim$ is naturally identified with $P^n(\boldsymbol{R})$.
4. Give a careful definition in terms of equivalence relations of the cutting and pasting (identification) procedures used to obtain examples of topological manifolds in Section I.4.
5. Use the ideas of Exercise 4 to show that $P^2(\boldsymbol{R})$ can be obtained by defining a suitable equivalence relation on a square.
6. Let X, Y be topological spaces, $\sim$ an equivalence relation on X and F: $X \to Y$ a continuous mapping. Suppose that $x_1 \sim x_2$ implies $F(x_1) = F(x_2)$ and show that F induces a continuous mapping $F': X/\sim \to Y$ such that $F' \circ \pi = F$.
7. Show that $\pi: F(k, n) \to G(k, n)$ is an open mapping and that $G(k, n)$ is Hausdorff. [*Hint:* show that $\mathbf{x} \sim \mathbf{y}$ if and only if a certain collection of $(k + 1) \times (k + 1)$ minor determinants of the $2k \times n$ matrix whose rows are $x_1, \ldots, x_k; y, \ldots, y_k$ all vanish and apply Lemma 2.4.]
8. In the case $G(2, 4)$ complete the proof that this space is a C^∞ manifold.
9. Let $J_1, \ldots, J_N$, $N = \binom{n}{k}$, be the collection of distinct subsets $(1, 2, \ldots, n)$ containing k integers and for each $\mathbf{x} \in F(k, n)$, let $|\mathbf{x}_J| = \det \mathbf{x}_J$. Define a mapping $\Phi: F(k, n) \to P^N(\boldsymbol{R})$ by $\Phi(\mathbf{x}) = [(|\mathbf{x}_{J_1}|, \ldots, |\mathbf{x}_{J_N}|)]$ and show $\mathbf{x} \sim \mathbf{y}$ implies $\Phi(\mathbf{x}) = \Phi(\mathbf{y})$ so that Φ defines a mapping of $G(k, n)$ into $P^N(\boldsymbol{R})$. Show that this mapping is continuous and univalent. Use this to prove that $G(k, n)$ is Hausdorff.

3 Differentiable Functions and Mappings

On a topological space the concept of continuity has meaning; in an analogous way, on a C^∞ manifold we may define the concept of C^∞ function.

Let f be a real-valued function defined on an open set W_f of a C^∞ manifold M, possibly all of M; in brief, f: $W_f \to \boldsymbol{R}$. If U, φ is a coordinate neighborhood such that $W_f \cap U \neq \varnothing$ and if $x^1, \ldots, x^n$ denotes the local coordinates, then f corresponds to a function $\hat{f}(x^1, \ldots, x^n)$ on $\varphi(W_f \cap U)$ defined by $\hat{f} = f \circ \varphi^{-1}$, that is, so that $f(p) = \hat{f}(x^1(p), \ldots, x^n(p)) = \hat{f}(\varphi(p))$ for all $p \in W_f \cap U$. We will customarily omit the caret and use the same letter "f" for f as defined on W_f and for $\hat{f}$, its expression in local coordinates. Ordinarily this will result in no confusion; if two coordinate neighborhoods U, φ and V, ψ are involved, we will use different letters for the coordinates, say $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$. Thus for $p \in W_f \cap U \cap V$ we have, omitting carets,

$$f(p) = f(x^1(p), \ldots, x^n(p)) = f(y^1(p), \ldots, y^n(p)),$$

the latter two f's denoting $\hat{f}$'s, or $f \circ \varphi^{-1}$ and $f \circ \psi^{-1}$, respectively, the expressions in local coordinates.

(3.1) Definition Using the notation above, $f\colon W_f \to \boldsymbol{R}$ is a C^∞ *function* if each $p \in W_f$ lies in a coordinate neighborhood U, φ such that $f \circ \varphi^{-1}(x^1, \ldots, x^n) = \hat{f}(x^1, \ldots, x^n)$ is C^∞ on $\varphi(W_f \cap U)$. [Clearly, a C^∞ function is continuous.]

Among the C^∞ functions on M are the n-coordinate functions $(x^1(q), \ldots, x^n(q))$ of a coordinate neighborhood U, φ. More precisely, if $\pi^i\colon \boldsymbol{R}^n \to \boldsymbol{R}$ is defined by $\pi^i(x^1, \ldots, x^n) = x^i$, these functions are defined by $x^i(q) = \pi^i \circ \varphi(q)$, and their expression in local coordinates, on $\varphi(U)$, by

$$\hat{x}^i(x^1, \ldots, x^n) = x^i(\varphi^{-1}(x^1, \ldots, x^n)) = \pi^i(x^1, \ldots, x^n) = x^i.$$

As mentioned above, the caret is usually omitted so we have the statement $x^i(x^1, \ldots, x^n) = x^i$, $i = 1, \ldots, n$,—somewhat confusing since the same letter is used for a function and its values.

It is a consequence of the definition that if f is C^∞ on W and $V \subset W$ is an open set, then $f \mid V$ is C^∞ on V. Moreover, if W is a union of open sets on each of which a real-valued function f is C^∞, then f is C^∞ on W. Using the C^∞ compatibility of coordinate neighborhoods, it is easily verified that if f is C^∞ on W and V, ψ is any coordinate neighborhood intersecting W, then $f \circ \psi^{-1}$ is C^∞ on the open set $\psi(V \cap W)$ in $\boldsymbol{R}^n$.

Just as in the case of $\boldsymbol{R}^n$ we proceed from definition of C^∞ function to definition of C^∞ mapping. Suppose that M and N are C^∞ manifolds, $W \subset M$ is an open subset, and $F\colon W \to N$ is a mapping, then we make the following definition.

(3.2) Definition F is a C^∞ *mapping* of W into N if for every $p \in W$ there exist coordinate neighborhoods U, φ of p and V, ψ of $F(p)$ with $F(U) \subset V$ such that $\psi \circ F \circ \varphi^{-1}\colon \varphi(U) \to \psi(V)$ is C^∞ in the sense of Section II.2.

More precisely, this means that $F \mid U\colon U \to V$ may be written in local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^m$ as a mapping from $\varphi(U)$ into $\psi(V)$ by

$$\hat{F}(x^1, \ldots, x^n) = (f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)),$$

[or simply $y^i = f^i(x)$, $i = 1, \ldots, m$] and each $f^i(x)$ is C^∞ on $\varphi(U)$. Note that C^∞ mapping is a more general notion than C^∞ function, the latter being a mapping to $N = \boldsymbol{R}$, which is, of course, the same as $\boldsymbol{R}^1$.

(3.3) Remark It is important to note that C^∞ mappings are continuous; that their restrictions to open subsets are C^∞; and that any mapping from an open subset $W \subset N$ into M, whose restriction to each of a collection of

open sets which cover W is C^∞, is necessarily C^∞ on W. As with Definition 3.1, the C^∞ compatibility of local coordinate neighborhoods, Corollary II.2.4, and the remarks above show easily that the property does not depend on any particular choice of coordinates. Similarly it follows from the same corollary that composition of C^∞ mappings is again a C^∞ mapping.

Many authors refer to C^∞ manifolds, functions, and mappings as *smooth*. From now on we shall refer to *differentiable* manifold, function and mapping although this is not very logical since we previously (in Chapter II) used this word in a much weaker sense than C^∞. One reason that C^∞ is a desirable differentiability class to use is that when we later take derivatives of C^∞ functions on manifolds, we obtain C^∞ functions—in the C^r case we would obtain C^{r-1} functions. Thus assuming infinite differentiability relieves us of many irritating concerns about order of differentiability. Of course, the same would be true for C^ω (real-analytic), but this is too restrictive for most purposes since we are unable to obtain important theorems of the following type.

(3.4) Theorem *Let F be a closed subset and K a compact subset of a C^∞ manifold M with $F \cap K = \varnothing$. Then there is a C^∞ function f defined on M with values in $[0, 1]$ which has the value $+1$ on K and 0 on F.*

Proof The proof of this theorem and that of the following corollary require a slight modification of Theorem II.5.1. This is left to the reader as an exercise. ∎

(3.5) Corollary *Let U be an open subset of a manifold M, suppose $p \in U$, and let f be a C^∞ function on U. Then there is a neighborhood V of p in U and a C^∞ function f^* on M such that $f^* = f$ on V and $f^* = 0$ outside of U.*

We conclude this section with a definition, an example, and some remarks on a basic problem referred to in Section 1.

(3.6) Definition A C^∞ mapping $F: M \to N$ between C^∞ manifolds is a *diffeomorphism* if it is a homeomorphism and F^{-1} is C^∞. M and N are *diffeomorphic* if there exists a diffeomorphism $F: M \to N$.

This extends the concept of diffeomorphism, previously defined for open subsets of $\boldsymbol{R}^n$ only, to arbitrary C^∞ manifolds. Diffeomorphism of manifolds is an equivalence relation since composition of C^∞ maps is C^∞ and composition of homeomorphisms is a homeomorphism. From this transitivity follows; reflexivity and symmetry are obvious from the definition. It is important that F^{-1}, as well as F, be C^∞ as the following example shows.

(3.7) Example Let $F: \boldsymbol{R} \to \boldsymbol{R}$ be defined by $F(t) = t^3$. Then F is C^∞ and a homeomorphism, but it is not a diffeomorphism since $F^{-1}(t) = t^{1/3}$ and this is not even of class C^1—let alone C^∞—at $t = 0$. This same example shows how it is possible to define two distinct C^∞ structures on $\boldsymbol{R}$. The first is the usual one defined by letting $U = \boldsymbol{R}$ and $\varphi: U \to \boldsymbol{R}$ be the identity map; this determines a C^∞ structure on $\boldsymbol{R}$ by Theorem 1.3. We may also consider the structure defined by the coordinate neighborhood V, ψ with $V = \boldsymbol{R}$ and $\psi: V \to \boldsymbol{R}$ defined by $\psi(t) = t^3$. Then $\varphi \circ \psi^{-1}(t) = t^{1/3}$ so that U, φ and V, ψ are not C^∞-compatible and hence not in the same differentiable structure. However, $\boldsymbol{R}$ with its first structure—the usual one—is diffeomorphic to $\tilde{\boldsymbol{R}}$, denoting $\boldsymbol{R}$ with its second structure, the diffeomorphism $F: \boldsymbol{R} \to \tilde{\boldsymbol{R}}$ being defined by $F(t) = t^{1/3}$ so that in local coordinates it is given by $\psi \circ F \circ \varphi^{-1}(t) = t$.

We have just seen, then, that two C^∞ manifolds with the same underlying topological manifold but incompatible C^∞ structures can still be diffeomorphic. A fundamental question is: Can the same manifold M or homeomorphic manifolds have C^∞ structures which are not diffeomorphic? This was an unsolved problem for many years, and it was finally settled by Milnor [4] who proved the existence of two C^∞ structures on S^7 which were not diffeomorphic. Further details are given in the notes.

We conclude with a remark which is occasionally useful: A necessary and sufficient condition that an open set U of M, together with a mapping $\varphi: U \to \boldsymbol{R}^n$, be a coordinate neighborhood is that φ be a diffeomorphism of U onto an open subset W of $\boldsymbol{R}^n$. Conversely, if W is an open subset of $\boldsymbol{R}^n$ and $\psi: W \to M$ is a diffeomorphism onto an open subset U, then U, ψ^{-1} is a coordinate neighborhood. We sometimes call W, ψ a *parametrization*, especially in the case $\dim M = 1$.

Exercises

1. Show that a continuous mapping $F: M \to N$, C^∞ manifolds, is C^∞ if and only if for any C^∞ function f on an open set $W_f \subset N$ the function $f \circ F$ is C^∞.
2. Verify the statements of Remark 3.3.
3. Prove the statement of the concluding paragraph of this section.
4. Assume that K is a compact subset of a C^∞ manifold M which is contained in an open subset U. Prove that there is a C^∞ function f: $M \to [0, 1]$ which is identically 1 on K and whose support (closure of the set on which $f \neq 0$) is contained in U.
5. Let M, N, and A be C^∞ manifolds and $p_1: M \times N \to M$, $p_2: M \times N \to N$ be projections to the factors. For $(a, b) \in M \times N$, let $i: M \to M \times N$ be defined by $i(p) = (p, b)$ and $j: N \to M \times N$ by $j(q) =$

(a, q). Show that p_1, p_2, i, and j are C^∞ mappings. Show that a mapping $F: A \to M \times N$ is C^∞ if and only if $f_1 = p_1 \circ F$ and $f_2 = p_2 \circ F$ are C^∞.

6. Suppose M and N are C^∞ manifolds, U an open set of M, and $F: U \to N$ is C^∞. Show that there exists a neighborhood V of any $p \in U$, $V \subset U$, such that F can be extended to a C^∞ mapping $F^*: M \to N$ with $F(q) = F^*(q)$ for all $q \in V$.
7. Identify the set of k-frames $F(k, n)$ of $\boldsymbol{R}^n$ with the set of $k \times n$ matrices of rank k in $\mathscr{M}_{kn}(\boldsymbol{R})$ and let $\pi: F(k, n) \to G(k, n)$ be the mapping taking each such matrix to its equivalence class (see Example 2.6). Show that this map is C^∞ relative to the differentiable structure of $F(k, n)$ as an open submanifold of $\mathscr{M}_{kn}(\boldsymbol{R})$.
8. Let A, B, M, N be C^∞ manifolds and let $F: A \to M$ and $G: B \to N$ be C^∞ mappings. Show that $F \times G: A \times B \to M \times N$ is C^∞, where $(F \times G)(x, y) = (F(x), G(y))$.

4 Rank of a Mapping. Immersions

Let $F: N \to M$ be a differentiable mapping of C^∞ manifolds and let $p \in N$. If U, φ and V, ψ are coordinate neighborhoods of p and $F(p)$, respectively, and $F(U) \subset V$, then we have a corresponding expression for F in local coordinates, namely,

$$\hat{F} = \psi \circ F \circ \varphi^{-1}: \varphi(U) \to \psi(V).$$

(4.1) Definition The *rank of F at p* is defined to be the rank of $\hat{F}$ at $\varphi(p)$ (as in Section II.8).

Thus the rank at p is the rank at $a = \varphi(p)$ of the Jacobian matrix

$$\begin{pmatrix} \dfrac{\partial f^1}{\partial x^1} & \cdots & \dfrac{\partial f^1}{\partial x^n} \\ \vdots & & \vdots \\ \dfrac{\partial f^m}{\partial x^1} & \cdots & \dfrac{\partial f^m}{\partial x^n} \end{pmatrix}_a$$

of the mapping $\hat{F}(x^1, \ldots, x^n) = (f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n))$ expressing F in the local coordinates. This definition must be validated by showing that the rank is independent of the choice of coordinates. This is left as an exercise—a second definition which is clearly independent of the choice is given in the next chapter.

(4.2) Remark As might be conjectured, the important case for us will be that in which the rank is constant. In fact Theorem II.7.1 (theorem on rank) and its Corollary II.7.2 can be restated as follows:

Let $F: N \to M$ be as above and suppose $\dim N = n$, $\dim M = m$ *and rank* $F = k$ *at every point of* N. *If* $p \in N$, *then there exist coordinate neighborhoods* U, φ *and* V, ψ *as above such that* $\varphi(p) = (0, \ldots, 0)$, $\psi(F(p)) = (0, \ldots, 0)$ *and* $\hat{F} = \psi \circ F \circ \varphi^{-1}$ *is given by*

$$\hat{F}(x^1, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0).$$

Moreover we may assume $\varphi(U) = C_\varepsilon^n(0)$ *and* $\psi(V) = C_\varepsilon^m(0)$ *with the same* $\varepsilon > 0$.

An obvious corollary to this remark is: a necessary condition for $F: N \to M$ to be a diffeomorphism is that $\dim M = \dim N = \operatorname{rank} F$. Otherwise k would be either less than n or less than m, in which case the expression in local coordinates implies that it is not possible for both F and F^{-1} to be one-to-one, even locally. For example, if $k < n$ in the expression above, all points in U with coordinates of the form $(0, \ldots, 0, x^{k+1}, \ldots, x^n)$ are mapped onto the same point of V.

(4.3) Definition $F: N \to M$ is said to be an *immersion* (*submersion*) if rank $F = n = \dim N (= m = \dim M)$ everywhere. If F is an injective (one-to-one) immersion, then F establishes a one-to-one correspondence of N and the subset $\tilde{N} = F(N)$ of M. If we use this correspondence to endow $\tilde{N}$ with a topology and C^∞ structure, then $\tilde{N}$ will be called a *submanifold* (or immersed submanifold) and $F: N \to \tilde{N}$ is a diffeomorphism.

We note that rank $F \leq \max(m, n)$ at every point. It follows that if F is an immersion (submersion), then $n \leq m$ (respectively, $n \geq m$). In general the topology and C^∞ structure of an immersed submanifold $\tilde{N} \subset M$ depend on F as well as N, i.e., $\tilde{N}$ is not a *subspace* of M. The concept of submanifold will be fully discussed below after considering a number of examples of immersions F of $\boldsymbol{R}$ or an open interval $\boldsymbol{R}$ into $\boldsymbol{R}^n$.

We use the natural coordinates t on $\boldsymbol{R}$ and (x^1, x^2) on $\boldsymbol{R}^2$ to define F. To verify that F is an immersion it is necessary to check that the Jacobian has rank 1 at every point, that is, that one of the derivatives with respect to t differs from zero for every value of t for which the mapping F is defined; this is left to the reader.

(4.4) Example $F: \boldsymbol{R} \to \boldsymbol{R}^3$ is given by $F(t) = (\cos 2\pi t, \sin 2\pi t, t)$. The image $F(\boldsymbol{R})$ is a *helix* lying on a unit cylinder whose axis is the x^3-axis in $\boldsymbol{R}^3$ (Fig. III.5a).

(4.5) Example $F: \boldsymbol{R} \to \boldsymbol{R}^2$ is given by $F(t) = (\cos 2\pi t, \sin 2\pi t)$. The image $F(\boldsymbol{R})$ is the unit circle $S^1 = \{(x^1, x^2) | (x^1)^2 + (x^2)^2 = 1\}$ in $\boldsymbol{R}^2$ (Fig. III.5b).

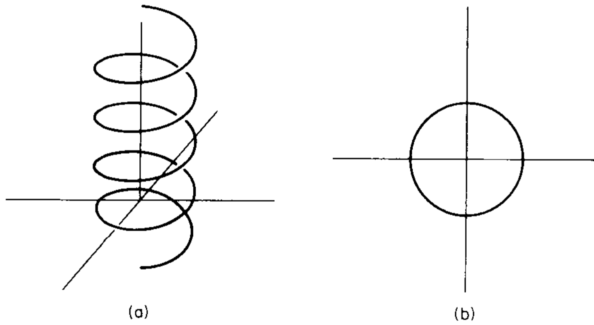


Figure III.5

(4.6) Example $F: (1, \infty) \to \boldsymbol{R}^2$ is given by

$$F(t) = ((1/t)\cos 2\pi t, (1/t)\sin 2\pi t).$$

The image is a curve spiraling to $(0, 0)$ as $t \to \infty$ and tending to $(1, 0)$ as $t \to 1$ (Fig. III.6a).

(4.7) Example $F: (1, \infty) \to \boldsymbol{R}^2$, as in the previous example. However, F is modified so that the image $F(\boldsymbol{R})$ spirals toward the circle with center at $(0, 0)$ and radius $\frac{1}{2}$ as $t \to \infty$. The mapping is given by

$$F(t) = \left(\frac{t+1}{2t}\cos 2\pi t, \frac{t+1}{2t}\sin 2\pi t\right).$$

[It is not difficult to check that the Jacobian could have rank 0, that is, both derivatives dx^1/dt and dx^2/dt could vanish simultaneously on $1 < t < \infty$ if and only if $\cot 2\pi t = -\tan 2\pi t$, which is impossible (Fig. III.6b).]

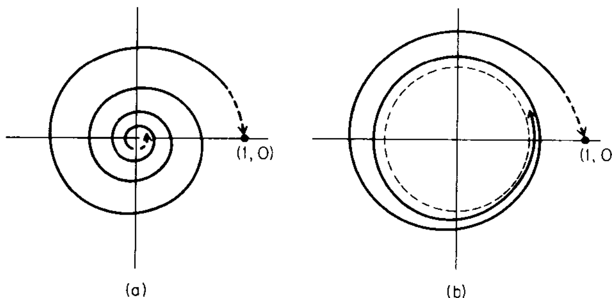


Figure III.6

(4.8) Example $F: \boldsymbol{R} \to \boldsymbol{R}^2$ is given by

$$F(t) = (2\cos(t - \tfrac{1}{2}\pi), \sin 2(t - \tfrac{1}{2}\pi)).$$

The image is a "figure eight" traversed in the sense shown (Fig. III.7a) with

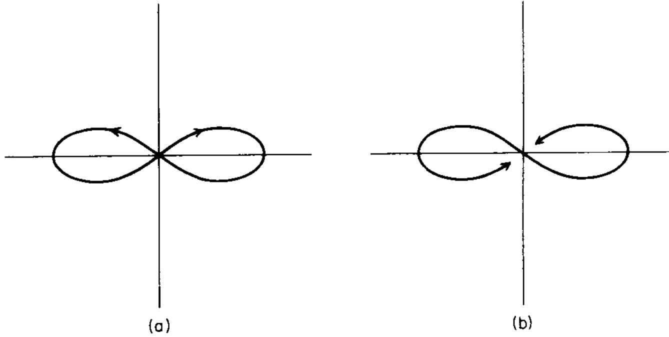


Figure III.7

the image point making a complete circuit starting at the origin as t goes from 0 to 2π.

(4.9) Example $G: \boldsymbol{R} \to \boldsymbol{R}^2$ again and the image is the "figure eight" as in the previous example, but with an important difference: we pass through $(0, 0)$ only once, when $t = \frac{1}{2}$. For $t \to -\infty$ and $t \to +\infty$ we only approach $(0, 0)$ as limit—as shown in Fig. III.7b. The immersion is given by changing parameter in the previous example: Let $g(t)$ be a monotone increasing C^∞ function on $-\infty < t < \infty$ such that $g(0) = \pi$, $\lim_{t \to -\infty} g(t) = 0$ and $\lim_{t \to +\infty} g(t) = 2\pi$. For example, we may use $g(t) = \pi + 2\tan^{-1} t$. Then $G(t)$ is given by composition of $g(t)$ with $F(t)$ from the previous example:

$$G(t) = F(g(t)) = \left(2\cos\left(g(t) - \frac{\pi}{2}\right), \sin 2\left(g(t) - \frac{\pi}{2}\right)\right).$$

(4.10) Example Again $F: \boldsymbol{R} \to \boldsymbol{R}^2$ so that

$$F(t) = \begin{cases} \left(\dfrac{1}{t}, \sin \pi t\right) & \text{for} \quad 1 \le t < \infty, \\ (0, t + 2) & \text{for} \quad -\infty < t \le -1. \end{cases}$$

This gives a curve with a gap as shown in Fig. III.8. For $-1 \le t \le +1$ we connect the two pieces together smoothly as shown by the dotted line. This gives a C^∞ immersion of all of $\boldsymbol{R}$ in $\boldsymbol{R}^2$ whose image is as shown. As we shall see, this is a useful example to keep in mind.

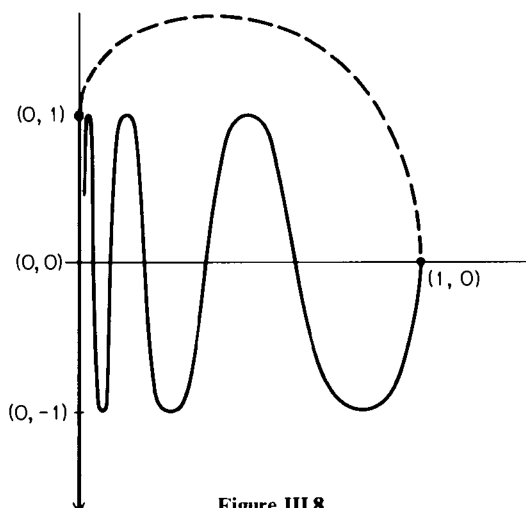


Figure III.8

We may draw some conclusions from these examples about the nature of immersions. First we note that an immersion need not be univalent, that is, one-to-one into (injective), in the large, even though it is one-to-one locally as we see from Remark 4.2. Examples 4.5 and 4.8 show this since, for example, in both cases $t = 0, \pm 2\pi, \pm 4\pi, \ldots$ all have the same image point: (0, 1) in the case of the circle and (0, 0) for the figure eight.

The second conclusion we can draw is that even when it is one-to-one, an immersion is not necessarily a homeomorphism onto its image, that is, $F: N \to M$ a one-to-one immersion does *not* imply that F is a homeomorphism of N onto $\tilde{N} = F(N)$ considered as a *subspace* of M. Examples 4.9 and 4.10 show this: in the case of Example 4.9, $\tilde{N}$ is the figure eight whereas N is the real line $\boldsymbol{R}$—two spaces which are not homeomorphic. In the case of Example 4.10, N is again the real line and $\tilde{N} = F(N)$ as a subspace of $\boldsymbol{R}^2$ is not locally connected at all of its points: there are points on the x^2-axis such as $(0, \frac{1}{2})$, which do not have arbitarily small connected neighborhoods; hence $\tilde{N}$ and $N = \boldsymbol{R}$ are not homeomorphic. In any case, of course, $F: N \to M$ is continuous—since it is differentiable. These examples lead to the definition of a more restrictive concept.

(4.11) Definition An *imbedding* is a one-to-one immersion $F: N \to M$ which is a homeomorphism of N into M, that is, F is a homeomorphism of N onto its image, $\tilde{N} = F(N)$, with its topology as a subspace of M. The image of an imbedding is called an *imbedded submanifold.*

We remark that Examples 4.4, 4.6, and 4.7 are imbeddings. The following theorem, essentially a restatement of the theorem on rank and its corollary

along the lines of a remark above, show that the distinction between immersions and imbeddings is a *global* one—it does not depend on the nature of F locally.

(4.12) Theorem *Let $F: N \to M$ be an immersion. Then each $p \in N$ has a neighborhood U such that $F \mid U$ is an imbedding of U in M.*

Proof According to Remark 4.2, we may choose cubical coordinate neighborhoods U, φ and V, ψ of $p \in N$ and $F(p) \in M$, respectively, such that $\varphi(p) = (0, \ldots, 0)$ in $\boldsymbol{R}^n$, $\psi(F(p)) = (0, \ldots, 0)$ in $\boldsymbol{R}^m$ with $\varphi(U) = C_\varepsilon^n(0)$ and $\psi(V) = C_\varepsilon^m(0)$ (cubes of the same breadth ε) and such that $\hat{F} = \psi \circ F \circ \varphi^{-1}$, the expression of F in these local coordinates, is given by

$$\hat{F}(x^1, \ldots, x^n) = (x^1, \ldots, x^n, 0, \ldots, 0).$$

To see that $F \mid U$ is a homeomorphism of U onto $F(U)$ with the relative topology, it is enough to see that $\hat{F}$ is a homeomorphism of $C_\varepsilon^n(0)$ onto its image in $C_\varepsilon^m(0)$. This is because $F(U) \subset V$, an open subset of M, so the topology of $F(U)$ as a subspace of M is the same as its topology as a subspace of V, and because $\varphi: U \to C_\varepsilon^n(0)$ and $\psi: V \to C_\varepsilon^m(0)$ are homeomorphisms. But it is clear that $\hat{F}$ is a homeomorphism of $C_\varepsilon^n(0)$ onto the subset $x^{n+1} = \cdots = x^m = 0$ of $C_\varepsilon^m(0)$; hence the theorem holds. ∎

(4.13) Remark It is convenient to call a subset S of a cube $C_\varepsilon^m(a)$ in $\boldsymbol{R}^m$ an n-slice if it consists of all points for which $m - n$ coordinates are held constant. For example, $S = \{x \in C_\varepsilon^m(0) \mid x^{n+1} = \cdots = x^m = 0\}$ is a slice through the center $0 = (0, \ldots, 0)$ of $C_\varepsilon^m(0)$. If V, ψ is a cubical coordinate neighborhood on a manifold M and S' is a subset of V such that $\psi(S')$ is a slice S of the cube $\psi(V)$, then S' is called a slice of V. It is an imbedded submanifold of M with imbedding $\psi^{-1}: S \to M$.

We note for future use that in the proof of Theorem 4.12, $S' = F(U)$ is a slice of V. In general this slice is not equal to the set $V \cap F(N)$ but only contained in it, even if F is univalent and U is chosen very small. The reader should verify this using the preceding examples.

Exercises

1. Using Exercise II.2.4, show that the rank of a mapping of C^∞ manifolds is independent of the choices of local coordinates made in Definition 4.1.
2. Show that a submersion is an open mapping and that if both manifolds have the same dimension, it is locally—but not necessarily globally—a diffeomorphism onto its image.
3. On the "figure eight" image $\tilde{N}$ of $\boldsymbol{R}$ in Example 4.9 let the topology and C^∞ structure be given by the one-to-one immersion G. Show that the

reflection $H: (x^1, x^2) \to (x^1, -x^2)$ of $\boldsymbol{R}^2$ in the x^1 axis, although it maps $\tilde{N}$ onto $\tilde{N}$, is not a diffeomorphism of $\tilde{N}$.

4. Let $F_i: N_i \to M$, $i = 1, 2$ be one-to-one immersions. Define them to be equivalent if there is a diffeomorphism $G: N_1 \to N_2$ such that $F_1 = F_2 \circ G$. Show that this is an equivalence relation (see Warner [1, Section 1.33]).
5. Show by example that there exist two inequivalent one-to-one immersions of $\boldsymbol{R}$ into $\boldsymbol{R}^2$ with the same image.
6. Show that composition of immersions is an immersion.
7. Show that the map $F: S^{n-1} \to P^{n-1}(\boldsymbol{R})$, defined by $F(x^1, \ldots, x^n) = [x^1, \ldots, x^n]$, is C^∞ and everywhere of rank $n - 1$.
8. Let $F: M \to B$ be a C^∞ mapping of manifolds and let A be an (immersed) submanifold of M. Show that $F \mid A$ is a C^∞ mapping into B.

5 Submanifolds

In this section we shall discuss in some detail the various types of submanifold. This term is used in more than one sense in the literature; however all agree that a submanifold N of a differentiable manifold M is a *subset* which is itself a differentiable manifold. The confusion arises over the question of whether or not it should be required to be a *subspace* of M, that is, to have the relative topology. We have adopted the definition which is the most general, namely, a *submanifold* N is the image in M of a one-to-one immersion $F: N' \to M$, $N = F(N')$, of a manifold N' into M together with the topology and C^∞ structure which makes $F: N' \to N$ a diffeomorphism. This definition, particularly convenient for Lie groups, became widely used following the influential Chevalley [1]. Its disadvantages are clear from Examples 4.9 and 4.10 and Exercises 4.3–4.5. A more natural notion—which we shall now develop, is that of a *regular submanifold*; as its name implies, it will be a special case of the one above. It is more natural since its topology and differentiable structure are derived directly and unambiguously from that of M. We will first state the characteristic feature of those subsets of a differentiable manifold M which are regular submanifolds; to do so we suppose $m = \dim M$ and that n is an integer, $0 \leq n \leq m$.

(5.1) Definition A subset N of a C^∞ manifold M is said to have the *n-submanifold property* if each $p \in N$ has a coordinate neighborhood U, φ on M with local coordinates $x^1, \ldots, x^m$ such that (i) $\varphi(p) = (0, \ldots, 0)$, (ii) $\varphi(U) = C_\varepsilon^m(0)$, and (iii) $\varphi(U \cap N) = \{x \in C_\varepsilon^m(0) \mid x^{n+1} = \cdots = x^m = 0\}$. If N has this property, coordinate neighborhoods of this type are called *preferred* coordinates (relative to N). Figure III.9 shows such a subset N in $M = \boldsymbol{R}^3$ ($n = 2$ and $m = 3$).

Note that immersed submanifolds do not always have this property, for example, take $p = (0, 0)$ in Examples 4.9 and 4.10, and take note of the second paragraph of Remark 4.13.

Denote by $\pi: \boldsymbol{R}^m \to \boldsymbol{R}^n$, $n \leq m$, the projection to the first n coordinates, then we may state the following lemma, using the notation above.

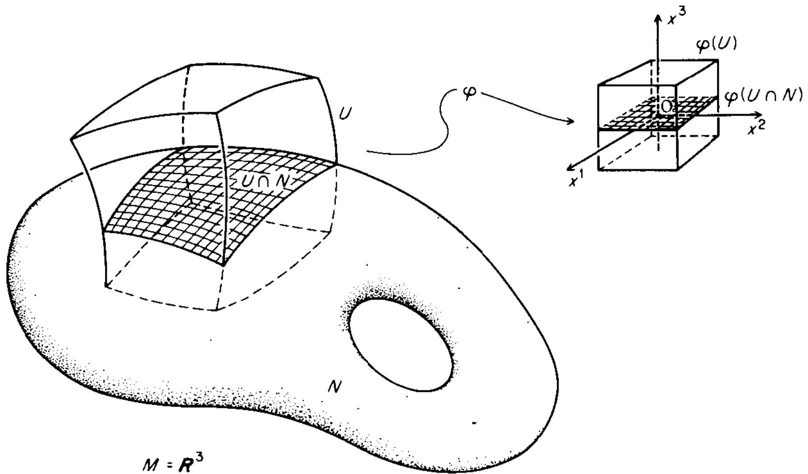


Figure III.9

(5.2) Lemma *Let $N \subset M$ have the n-submanifold property. Then N with the relative topology is a topological n manifold and each preferred coordinate system U, φ of M (relative to N) defines a local coordinate neighborhood V, $\tilde{\varphi}$ on N by $V = U \cap N$ and $\tilde{\varphi} = \pi \circ \varphi \mid V$. These local coordinates on N are C^∞-compatible wherever they overlap and determine a C^∞ structure on N relative to which the inclusion $i: N \to M$ is an imbedding.*

Proof Assume N has the subspace topology relative to M. Then V, $\tilde{\varphi}$ are topological coordinate neighborhoods covering N; for $V = U \cap N$ is an open set in the relative topology and $\tilde{\varphi}$ is a homeomorphism onto $C_\varepsilon^n(0) = \pi(C_\varepsilon^m(0))$ in $\boldsymbol{R}^n$. Suppose that for two preferred neighborhoods, U, φ and U', φ', $V = U \cap N$ and $V' = U' \cap N$ have nonempty intersection. Since V, $\tilde{\varphi}$ and V', $\tilde{\varphi}'$ are topological coordinate neighborhoods, we know that the change of coordinates is given by homeomorphisms $\tilde{\varphi}' \circ \tilde{\varphi}^{-1}$ and $\tilde{\varphi} \circ (\tilde{\varphi}')^{-1}$, which we must show to be C^∞. Let $\theta: \boldsymbol{R}^n \to \boldsymbol{R}^m$ be given by $\theta(x^1, \ldots, x^n) = (x^1, \ldots, x^n, 0, \ldots, 0)$ so that $\pi \circ \theta$ is the identity on $\boldsymbol{R}^n$. This map θ is C^∞ as is its restriction to $C_\varepsilon^n(0)$, an open subset of $\boldsymbol{R}^n$; thus $\tilde{\varphi}^{-1} = \varphi^{-1} \circ \theta$ is C^∞ since it is a composition of C^∞ maps. On the other hand $\tilde{\varphi}' = \pi \circ \varphi'$, and because φ' is a C^∞ map of U' and its open subset $U' \cap U$ to $\boldsymbol{R}^m$ and π is open and C^∞, we see that $\tilde{\varphi}'$ is C^∞ on $V \cap V'$. Thus $\tilde{\varphi}' \circ \tilde{\varphi}^{-1}$ is C^∞ on its domain, $\tilde{\varphi}(V \cap V')$.

It is even easier to see this if we write the expressions in local coordinates. If $y^i = f^i(x^1, \ldots, x^m)$, $i = 1, \ldots, m$ are the functions giving $\varphi' \circ \varphi^{-1}$, which we know to be C^∞, then it is easily checked that $\tilde{\varphi}' \circ \tilde{\varphi}^{-1}$ is given by $y^i = f^i(x^1, \ldots, x^n, 0, \ldots, 0)$, $i = 1, \ldots, n$. Therefore $\tilde{\varphi}' \circ \tilde{\varphi}^{-1}$ is C^∞ by Definition 3.2.

By Theorem 1.3 of this chapter, the totality of these neighborhoods determines a differentiable structure on N. In preferred local coordinates $V, \tilde{\varphi}, i \colon N \to M$ is given on V by $(x^1, \ldots, x^n) \to (x^1, \ldots, x^n, 0, \ldots, 0)$, so it is obviously an immersion. Because we have taken the relative topology on N, $i \colon N \to M$ is by definition a homeomorphism to its image $i(N) = N$, with the subspace topology, that is, i is an imbedding. ∎

The foregoing completes the proof of Lemma 5.2 and allows us to make the following important definition.

(5.3) Definition A *regular submanifold* of a C^∞ manifold M is any subspace N with the submanifold property and with the C^∞ structure that the corresponding preferred coordinate neighborhoods determine on it. (Note: N is trivially a submanifold, as previously defined, since i is an imbedding.)

As an example we shall see that $S^2 = \{x \in \boldsymbol{R}^3 \mid \|x\| = 1\}$ is really a submanifold, as was indicated at the end of Section 1. If $q = (a^1, a^2, a^3)$ is an arbitrary point on S^2, it cannot lie on more than one coordinate axis. For convenience we suppose that it does not lie on the x^3-axis. We introduce the usual spherical coordinates (r, θ, φ); they are defined on $\boldsymbol{R}^3 - \{x^3\text{-axis}\}$ and if $(1, \theta_0, \varphi_0)$ are the coordinates of q we may change the coordinate map slightly so that r is replaced by $\tilde{r} = r - 1$, θ by $\tilde{\theta} = \theta - \theta_0$, and φ by $\tilde{\varphi} = \varphi - \varphi_0$. Then for sufficiently small ε, the neighborhood V, ψ with coordinate function $\psi \colon p \to (\tilde{r}(p), \tilde{\theta}(p), \tilde{\varphi}(p))$ defined for p such that $|\tilde{r}| < \varepsilon$, $|\tilde{\theta}| < \varepsilon$, and $|\tilde{\varphi}| < \varepsilon$ defines a coordinate neighborhood of q, with q having coordinates $(0, 0, 0)$ and with $V \cap S^2$ the open subset of S^2 corresponding to $\tilde{r} = 0$. The fact that these neighborhoods are compatible with the ones previously defined for S^2 (Example 1.8) can be proved by writing down the standard equations giving rectangular Cartesian coordinates as functions of the spherical coordinates.

(5.4) Remark At this point we have apparently defined three classes of submanifolds in a manifold M—*immersed*, *imbedded*, and *regular*. The first of these, which we usually call simply a *submanifold*, includes the last two as special cases. It was defined (4.3) as the image $N = F(N')$ of a C^∞ manifold N' under a univalent immersion F, and was by definition endowed with the C^∞ manifold structure of N'. This means simply that we use the one-to-one onto correspondence $F \colon N' \to N$ to topologize N, and define the C^∞ structure on N in terms of coordinate neighborhoods U, φ, where $U = F(U')$ and

$\varphi = (F|N)^{-1} \circ \varphi'$, using all coordinate neighborhoods U', φ' of the C^∞ structure of N'. A priori there is little relation between the topology and differentiable structure thus given to the subset N of M and that of M. However, the fact that F is continuous implies that the topology of N obtained in this way is in general finer than its relative topology as a subspace of M, that is, if V is open in M, then $V \cap N$ is open in N, but there may well be open sets of N which are not of this form.

An *imbedding* on the other hand is a particular type of univalent immersion, one in which U' is open in N' if and only if $F(U') = V \cap N$ for some open set V of M so that the topology of the submanifold $N = F(N')$ is exactly its relative topology as a subspace of M. An imbedded submanifold is thus a special type of (immersed) submanifold.

Finally, a *regular* submanifold is an imbedded submanifold with the inclusion i as imbedding map, and, conversely, as the next theorem shows, the submanifold structure on $N = F(N')$ given by any imbedding F is that of a regular submanifold. Hence we have (i) (immersed) submanifolds and (ii) their subclass imbedded or regular submanifolds, the latter being the most natural and most important. Perfectly legitimate alternative definitions would be as follows: let N be a subset of a C^∞ manifold M which is (somehow) independently given the topology and C^∞ structure of a manifold. Then N is a submanifold if for this structure the inclusion $i: N \to M$ is an immersion and a regular submanifold if i is an imbedding, i.e., its given topology is the relative topology as a subspace of M. Moreover,

(5.5) Theorem *Let $F: N' \to M$ be an imbedding of a C^∞ manifold N of dimension n in a C^∞ manifold M of dimension m. Then $N = F(N')$ has the n-submanifold property and thus N is a regular submanifold. As such it is diffeomorphic to N' with respect to the mapping $F: N' \to N$.*

Proof Let $q = F(p)$ be any point of N. As in Theorem 4.12, we choose cubical coordinate neighborhoods U, φ of p and V, ψ of q such that (i) $\varphi(p) = (0, \ldots, 0) \in C_\varepsilon^n(0) = \varphi(U)$, (ii) $\psi(q) = (0, \ldots, 0) \in C_\varepsilon^m(0) = \psi(V)$, and (iii) the mapping $F|U$ is given in local coordinates by $\hat{F}: (x^1, \ldots, x^n) \mapsto (x^1, \ldots, x^n, 0, \ldots, 0)$. If $F(U) = V \cap N$, then the neighborhood V, ψ would be a preferred coordinate neighborhood relative to N. We may show this to hold for restrictions $U' \subset U$ and $V' \subset V$. Since F is an imbedding, $F(U)$ is a relatively open set of N, that is, $F(U) = W \cap N$, where W is open in M, and since $V \supset F(U)$, it is no loss of generality to suppose $W \subset V$. Thus $\psi(W)$ is an open subset of $C_\varepsilon^m(0)$ containing the origin and $\psi(W) \supset \psi(F(U))$, which is a slice S of $C_\varepsilon^m(0)$, $S = \{x \in C_\varepsilon^m(0) | x^{n+1} = \cdots = x^m = 0\}$. Therefore we may choose a (smaller) open cube $C_\delta^m(0) \subset \psi(W)$ and let $V' = \psi^{-1}(C_\delta^m(0))$, $\psi' = \psi|V'$. This is a cubical coordinate neighborhood of q for which $F(U) \cap V' = V' \cap N$; moreover, taking $U' = \varphi^{-1}(C_\delta^n(0)) = F^{-1}(V')$, we see that U', φ', with $\varphi' = \varphi|U'$, is a coordinate neighborhood of p and the pair U', φ',

and V', ψ' have exactly the properties needed, namely, (i), (ii), (iii) and $F(U') = V' \cap N$. This proves simultaneously that N has the n-submanifold property and that F is a diffeomorphism. The latter is true since the inverse of $F: N' \to N$ is given in the local preferred coordinates V', $\pi \circ \psi'$ and U', φ' by $\hat{F}^{-1}(x^1, \ldots, x^n) = (x^1, \ldots, x^n)$, which is clearly C^∞. ∎

(5.6) Remark The following addendum to Remark 4.13 and to the preceding comments is often useful. Suppose that $N \subset M$ is an (immersed) submanifold and that $q \in N$. Then there is a cubical neighborhood V, ψ of q with $\psi(q) = (0, \ldots, 0) \in C_\varepsilon^m(0) = \psi(V)$ such that the slice $S' \subset V$, consisting of all points of V whose last $m - n$ coordinates vanish, is an open set and a cubical coordinate neighborhood U, φ of the submanifold structure of N with $U = S'$ and $\varphi(r) = \pi \circ \psi(r) = x^1(r), \ldots, x^n(r))$. Thus the coordinates on U are just the restrictions to U of the first n coordinates of V, ψ. The proof is left as an exercise.

We now have a definition of regular submanifold, and we wish to obtain examples—which are useful for many reasons, among which is that they give further interesting examples of manifolds. Since it is usually easier to determine that a map from one C^∞ manifold into another is an immersion than to see that it is an imbedding, the following theorem is useful. A generalization is given in Exercise 1.

(5.7) Theorem *If $F: N \to M$ is a one-to-one immersion and N is compact, then F is an imbedding and $\tilde{N} = F(N)$ a regular submanifold. Thus a submanifold of M, if compact in M, is regular.*

Proof Since F is continuous and both N and $\tilde{N}$—with the subspace topology—are Hausdorff, we have a continuous (one-to-one) mapping from a compact space to a Hausdorff space. Since a closed subset K of N is compact, $F(K)$ is compact and therefore closed. Thus F takes closed subsets of N to closed subsets of $\tilde{N}$, and being one-to-one onto it takes open subsets to open subsets also. It follows that F^{-1} is continuous so $F: N \to \tilde{N}$ is a homeomorphism and therefore an imbedding. The rest of the statement follows from our remarks above. ∎

The most useful method of finding examples of submanifolds is given by the following theorem and its corollary. Since many examples of manifolds, as we have seen, occur as submanifolds of some other manifold, especially Euclidean space, the corollary is also very helpful in proving that some of the objects we have looked at are indeed C^∞ manifolds. Examples are given below.

(5.8) Theorem *Let N be a C^∞ manifold of dimension n, M be a C^∞ manifold of dimension m, and $F: N \to M$ be a C^∞ mapping. Suppose that F has constant*

rank k on N and that $q \in F(N)$. Then $F^{-1}(q)$ is a closed, regular submanifold of N of dimension $n - k$.

Proof Let A denote $F^{-1}(q)$; A is a closed subset since the inverse image of $\{q\}$, a closed subset of M, under a continuous map is closed. We shall show that A has the submanifold property for the dimension $n - k$. Let $p \in A$; since F has constant rank k on a neighborhood of p, by the theorem on rank (Remark 4.2) we may find coordinate neighborhoods U, φ and V, ψ of p and q, respectively, such that $\varphi(p)$ and $\psi(q)$ are the origins in $\boldsymbol{R}^n$ and $\boldsymbol{R}^m$, $\varphi(U) = C_\varepsilon^n(0)$, $\psi(V) = C_\varepsilon^m(0)$, and in local coordinate $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^m)$, $F \mid U$ is given by the mapping

$$\psi \circ F \circ \varphi^{-1}(x) = \hat{F}(x^1, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0).$$

This means that the only points of U mapping onto q are those whose first k coordinates are zero, that is,

$$\begin{aligned} A \cap U &= \varphi^{-1}(\varphi \circ F^{-1} \circ \psi^{-1}(0)) \\ &= \varphi^{-1}(\hat{F}^{-1}(0)) = \varphi^{-1}\{x \in C_\varepsilon^n(0) \mid x^1 = \cdots = x^k = 0\}. \end{aligned}$$

Hence A is a regular manifold of dimension $n - k$ since it has the submanifold property. ∎

(5.9) Corollary *If $F: N \to M$ is a C^∞ mapping of manifolds,* dim $M = m \le n =$ dim N, *and if the rank of $F = m$ at every point of $A = F^{-1}(a)$, then A is a closed, regular submanifold of N.*

The corollary holds because at $p \in A$, F has the maximum rank possible, namely m. It follows from Section II.7 and the independence of rank on local coordinates that, in some neighborhood of p in N, F has this rank also; thus the rank of F is m on an open subset of N containing A. But such a subset is itself a manifold of dimension n—an open submanifold—to which we may apply the theorem.

The first two applications of Corollary 5.9 are just very simple ways of demonstrating that S^{n-1} and the torus, described, as in Example I.3.4, by rotating a circle around a line in its plane which does not intersect it, are both manifolds. Other applications will be made in the next section.

(5.10) Example The map $F: \boldsymbol{R}^n \to \boldsymbol{R}$ defined by $F(x^1, \ldots, x^n) = \sum_{i=1}^n (x^i)^2$ has rank 1 on $\boldsymbol{R}^n - \{0\}$, which contains $F^{-1}(+1) = S^{n-1}$. Thus S^{n-1} is an $(n - 1)$-dimensional submanifold of $\boldsymbol{R}^n$ by Corollary 5.9.

(5.11) Example The map $F: \boldsymbol{R}^3 \to \boldsymbol{R}$ given by

$$F(x^1, x^2, x^3) = (a - ((x^1)^2 + (x^2)^2)^{1/2})^2 + (x^3)^2$$

has rank 1 at each point of $F^{-1}(b^2)$, $a > b > 0$. Thus the locus $F^{-1}(b^2)$, the torus in $\boldsymbol{R}^3$, Fig. I.2b, is a submanifold.

Exercises

1. Let $F: N \to M$ be a one-to-one immersion which is *proper*, i.e., the inverse image of any compact set is compact. Show that F is an imbedding and that its image is a closed regular submanifold of M and conversely.
2. Show that the mappings $i: M \to M \times N$ and $j: N \to M \times N$ defined in Exercise 3.5 are imbeddings.
3. Let $A \subset M$ and $B \subset N$ be submanifolds of M and N. Show that $A \times B$ is (by inclusion) a submanifold of $M \times N$. Show that if A and B are regular, so is $A \times B$.
4. Prove the statement of Remark 5.6.
5. If N is a submanifold of M and V is an open subset of M, possibly connected, then show that $N \cap V$ is the union of a countable collection of connected open subsets of N (with its submanifold topology).
6. Show by example that there may be functions that are C^∞ on a submanifold N of M that cannot be obtained by restriction of a C^∞ function on M.
7. Let $S^2 = \{x \in \boldsymbol{R}^3 \mid \|x\| = 1\}$ and obtain $P^2(\boldsymbol{R})$ by indentification of antipodal points of S^2. Define $F: \boldsymbol{R}^3 \to \boldsymbol{R}^4$ by $F(x^1, x^2, x^3) = ((x^1)^2 - (x^2)^2, x^1x^2, x^1x^3, x^2x^3)$. Since $F(-x) = F(x)$, $F|S^2$ defines a mapping $\tilde{F}: P^2(\boldsymbol{R}) \to \boldsymbol{R}^4$. Show that $\tilde{F}$ is a C^∞ imbedding (cf. Hilbert and Cohn-Vossen [1]).

6 Lie Groups

The space $\boldsymbol{R}^n$ is a C^∞ manifold and at the same time an Abelian group with group operation given by componentwise addition. Moreover the algebraic and differentiable structures are related: $(x, y) \to x + y$ is a C^∞ mapping of the product manifold $\boldsymbol{R}^n \times \boldsymbol{R}^n$ onto $\boldsymbol{R}^n$, that is, the group operation is differentiable. We also see that the mapping of $\boldsymbol{R}^n$ onto $\boldsymbol{R}^n$ given by taking each element x to its inverse $-x$ is differentiable.

Now let G be a group which is at the same time a differentiable manifold. For $x, y \in G$ let xy denote their product and x^{-1} the inverse of x.

(6.1) Definition G is a *Lie group* provided that the mapping of $G \times G \to G$ defined by $(x, y) \to xy$ and the mapping of $G \to G$ defined by $x \to x^{-1}$ are both C^∞ mappings.

(6.2) Example $Gl(n, \boldsymbol{R})$, the set of nonsingular $n \times n$ matrices, is as we have seen, an open submanifold of $\mathscr{M}_n(\boldsymbol{R})$, the set of $n \times n$ real matrices

identified with $\boldsymbol{R}^{n^2}$. Moreover $Gl(n, \boldsymbol{R})$ is a group with respect to matrix multiplication. In fact, an $n \times n$ matrix A is nonsingular if and only if $\det A \neq 0$; but $\det(AB) = (\det A)(\det B)$ so if A and B are nonsingular, AB is also. An $n \times n$ matrix A is nonsingular, that is, $\det A \neq 0$, if and only if A has a multiplicative inverse; thus $Gl(n, \boldsymbol{R})$ is a group. Both the maps $(A, B) \to AB$ and $A \to A^{-1}$ are C^ω. The product has entries which are polynomials in the entries of A and B, and these entries are exactly the expressions in local coordinates of the product map, which is thus C^ω, hence C^∞. The inverse of $A = (a_{ij})$ may be written as $A^{-1} = (1/\det A)(\tilde{a}_{ij})$, where the $\tilde{a}_{ij}$ are the cofactors of A (thus polynomials in the entries of A) and where $\det A$ is a polynomial in these entries which does not vanish on $Gl(n, \boldsymbol{R})$. It follows that the entries of A^{-1} are rational functions on $Gl(n, \boldsymbol{R})$ with nonvanishing denominators, hence C^ω (and C^∞). Therefore $Gl(n, \boldsymbol{R})$ is a Lie group. A special case is $Gl(1, \boldsymbol{R}) = \boldsymbol{R}^*$, the multiplicative group of nonzero real numbers.

(6.3) Example Let $\boldsymbol{C}^*$ be the nonzero complex numbers. Then $\boldsymbol{C}^*$ is a group with respect to multiplication of complex numbers, the inverse being $z^{-1} = 1/z$. Moreover $\boldsymbol{C}^*$ is a one-dimensional C^∞ manifold covered by a single coordinate neighborhood $U = \boldsymbol{C}^*$ with coordinate map $z \to \varphi(z)$ given by $\varphi(x + iy) = (x, y)$ for $z = x + iy$. Using these coordinates, the product $w = zz'$, $z = x + iy$, and $z' = x' + iy'$ is given by

$$((x, y)(x', y')) \to (xx' - yy', xy' + yx')$$

and the mapping $z \to z^{-1}$ by

$$(x, y) \to \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2}\right).$$

This means that the two mappings of Definition 6.1 are C^∞; therefore $\boldsymbol{C}^*$ is a Lie group.

(6.4) Theorem *If G_1 and G_2 are Lie groups, then the direct product $G_1 \times G_2$ of these groups with the C^∞ structure of the Cartesian product of manifolds is a Lie group.*

The proof is left as an exercise.

(6.5) Example (The toral groups) The circle S^1 may be identified with the complex numbers of absolute value $+1$. Since $|z_1 z_2| = |z_1||z_2|$, it is a group with respect to multiplication of complex numbers—a subgroup of $\boldsymbol{C}^*$. It is a Lie group as can be checked directly or proved as a consequence of Example 6.3 and the next theorem. Combining this with Theorem 6.4, we see that $T^n = S^1 \times \cdots \times S^1$, the n-fold Cartesian product, is a Lie group. It is called the *toral group*. Since S^1 is abelian, T^n is Abelian also.

As might be expected, the subgroups—in the sense of algebra—of a Lie group which are also submanifolds play a special role.

(6.6) Theorem *Let G be a Lie group and H a subgroup which is also a regular submanifold. Then with its differentiable structure as a submanifold H is a Lie group.*

The proof requires the following lemma, which is important in its own right.

(6.7) Lemma *Let $F: A \to M$ be a C^∞ mapping of C^∞ manifolds and suppose $F(A) \subset N$, N being a regular submanifold of M. Then F is C^∞ as a mapping into N.*

Proof Since N is a regular submanifold of M, each point is contained in a preferred coordinate neighborhood. Let $p \in A$, let $q = F(p)$ be its image, and let U, φ be a neighborhood of p which maps into a preferred coordinate neighborhood V, ψ of q. We have $\psi(V) = C_\varepsilon^m(0)$ with $\psi(q) = (0, \ldots, 0)$, the origin of $\boldsymbol{R}^m$, $m = \dim M$; and $V \cap N$ consists of those points of V whose last $m - n$ coordinates are zero, $n = \dim N$. Let $(x^1, \ldots, x^p)$ be the local coordinates in U, φ on A. Then the expression in local coordinates for F is

$$\hat{F}(x^1, \ldots, x^p) = (f^1(x), \ldots, f^n(x), 0, \ldots, 0),$$

that is, $f^{n+1}(x) = \cdots = f^m(x) = 0$ since $F(A) \subset N$.

However, $V \cap N$, $\pi \circ \psi|_{V \cap N}$ is a coordinate neighborhood of q on N, so F, considered as a *mapping into* N, is given in local coordinates by

$$(x^1, \ldots, x^n) \to (f^1(x), \ldots, f^n(x)).$$

This is $\hat{F}$ followed by the projection π to the first n coordinates (projection of $\boldsymbol{R}^m$ to $\boldsymbol{R}^n$), which is a composition of C^∞ maps and is therefore C^∞.

(6.8) Remark Lemma 6.7 does not hold for general submanifolds! In Example 4.9 if we map the open interval $(-1, 1)$ by a mapping G into $N = F(\boldsymbol{R})$, the figure eight, so that it crosses the origin as shown in Fig.III.10, then G is C^∞ as a mapping into $\boldsymbol{R}^2$, but not even continuous as a mapping to N. Thus N is diffeomorphic to the real line by $F: \boldsymbol{R} \to N$, and identifying N and $\boldsymbol{R}$, we may think of G as taking part of the open interval $(-1, 1)$, say $(0, 1)$, onto the real numbers $t > 1$, 0 onto 0, and $(-1, 0)$, the remaining part, onto the real numbers $t < -1$. The image is not even connected, so G is not continuous. But this lemma *can* be generalized to all submanifolds by adding a continuity condition: see Exercise 5 (also Exercise 7 for an important consequence).

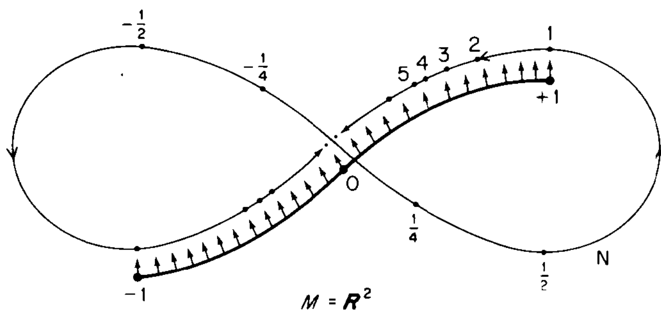


Figure III.10

Proof of the Theorem It follows without difficulty that $H \times H$ is a regular submanifold of $G \times G$. Thus the inclusion map $F_1 : H \times H \to G \times G$ is a C^∞ imbedding. If $F_2 : G \times G \to G$ is the C^∞ mapping $(g, g') \to gg'$ and $F = F_2 \circ F_1$ the composition, then F is a C^∞ mapping from $H \times H \to G$ with image in H. Let $\tilde{F}$ denote this map considered as a map into H; it is not the same mapping as F since we have changed the range. We must show that $\tilde{F}$ is C^∞ and similarly that the map $H \to G$ given by taking $h \to h^{-1}$ is C^∞ as a map onto H. These facts both follow from the lemma, which completes the proof. ∎

We may use Theorem 6.6 and Theorem 5.8 to give many further examples of Lie groups. To do so we make use of the following naturally defined maps of a Lie group G onto itself: (i) $x \to x^{-1}$, (ii) left and right translations by a fixed element a of G, that is, $L_a, R_a : G \to G$ defined by $L_a(x) = ax$ and $R_a(x) = xa$. These maps are C^∞ by definition of Lie group and have inverses which are C^∞, so they are, in fact, diffeomorphisms. The mapping of (i) is its own inverse and we have $(L_a)^{-1} = L_{a^{-1}}$ and $(R_a)^{-1} = R_{a^{-1}}$.

(6.9) Example $Sl(n, \boldsymbol{R}) = \{X \in Gl(n, \boldsymbol{R}) \mid \det X = +1\}$ is a subgroup and regular submanifold of $Gl(n, \boldsymbol{R})$, hence a Lie group. To prove this, we consider the mapping $F : Gl(n, \boldsymbol{R}) \to \boldsymbol{R}^*$, $F(X) = \det X$. According to the product rule, $\det(XY) = (\det X)(\det Y)$. Thus F is a homomorphism onto $\boldsymbol{R}^* = Gl(1, \boldsymbol{R})$; it is also C^∞ since it is given by polynomials in the entries. Finally, its rank is constant: Let $A \in Gl(n, \boldsymbol{R})$; let $a = \det A$; and let L_X, L_x denote left translations in $Gl(n, \boldsymbol{R})$ and $Gl(1, \boldsymbol{R}) = \boldsymbol{R}^*$. Then

$$F(X) = L_a \circ F \circ L_{A^{-1}}(X)$$

since we have $a \cdot \det(A^{-1}X) = \det X$. Applying the chain rule, Section II.2.3, using $DL_a = a \neq 0$, and using the fact that $L_{A^{-1}}$ is a diffeomorphism (so that $DL_{A^{-1}}$ is nonsingular), we have

$$\operatorname{rank} DF(X) = \operatorname{rank}[aDF(A^{-1}X)DL_{A^{-1}}(X)] = \operatorname{rank} DF(A^{-1}X)$$

for all $A \in Gl(n, \boldsymbol{R})$. In particular, rank $DF(X) =$ rank $DF(X^{-1}X) =$ rank $DF(I)$, and thus we see that the rank is constant as claimed. It follows that $Sl(n, \boldsymbol{R}) = F^{-1}(+1)$ is a closed, regular submanifold by Theorem 5.8. It is also a subgroup—in fact the kernel of a homomorphism—by virtue of the product rule for determinants; therefore it is a Lie group.

(6.10) Example $O(n) = \{X \in Gl(n, \boldsymbol{R}) \mid {}^tXX = I\}$, the subgroup of orthogonal $n \times n$ matrices is a regular submanifold and thus a Lie group. Let $F(X) = {}^tXX$, ${}^tX =$ transpose of X, define a mapping from $Gl(n, \boldsymbol{R})$ to $Gl(n, \boldsymbol{R})$. If $A \in Gl(n, \boldsymbol{R})$, we will show that rank $DF(X) =$ rank $DF(XA^{-1})$; and since any $Y \in Gl(n, \boldsymbol{R})$ can be written in the form $Y = XA^{-1}$, it follows that rank DF is constant on $Gl(n, \boldsymbol{R})$. To obtain this equality we note that $F(XA^{-1}) = L_{{}^tA^{-1}} \circ R_{A^{-1}} \circ F(X)$. Therefore

$$DF(XA^{-1}) = DL_{{}^tA^{-1}} \circ DR_{A^{-1}} \circ DF(X),$$

where $DR_{A^{-1}}$ and $DL_{{}^tA^{-1}}$ are evaluated at $F(X)$ and $R_{A^{-1}}(F(X))$, respectively. Then the equality of rank $DF(XA^{-1})$ and rank $DF(X)$ follows as above from the fact that $DL_{{}^tA^{-1}}$ and $DR_{A^{-1}}$ are everywhere nonsingular. Since $O(n) = F^{-1}(I)$, where I is the identity matrix, the statement follows from Theorem 5.8.

(6.11) Definition Let $F: G_1 \to G_2$ be an algebraic homomorphism of Lie groups G_1 and G_2. We shall call F a *homomorphism* (of Lie groups) if F is also a C^∞ mapping.

(6.12) Example Let $G_1 = Gl(n, \boldsymbol{R})$ and $G_2 = \boldsymbol{R}^*$ $[= Gl(1, \boldsymbol{R})]$. Then the map F given by $F(X) = \det X$ is a homomorphism.

(6.13) Example Let $G_1 = \boldsymbol{R}$, the additive group of real numbers, and $G_2 = S^1$, identified with the multiplicative group of complex numbers of absolute value 1. Then the mapping $F(t) = e^{2\pi i t}$ is a homomorphism. Similarly, letting $G_1 = \boldsymbol{R}^n$, a Lie group with componentwise addition, and $G_2 = T^n = S^1 \times \cdots \times S^1$, the mapping $F: \boldsymbol{R}^n \to T^n$ given by $F(t_1, \ldots, t_n) = (\exp 2\pi i t_1, \ldots, \exp 2\pi i t_n)$ is a homomorphism. Its kernel is the discrete additive group $\boldsymbol{Z}^n$ consisting of all n-tuples of integers; it is called the *integral lattice* of $\boldsymbol{R}^n$.

(6.14) Theorem *If $F: G_1 \to G_2$ is a homomorphism of Lie groups, then the rank of F is constant; the kernel is a closed regular submanifold and thus a Lie group; and* $\dim \ker F = \dim G_1 - \operatorname{rank} F$.

Proof Let $a \in G_1$ be arbitrarily chosen and let $b = F(a)$ be its image in G_2. Denote by e_1, e_2 the unit elements of G_1, G_2, respectively. Then we may write

$$F(x) = F(aa^{-1}x) = F(a)F(a^{-1}x) = L_b \circ F(a^{-1}x),$$

so that for all $x \in G_1$

$$DF(x) = DL_b \cdot DF(a^{-1}x).$$

Then, since L_b is a diffeomorphism, and thus has nonsingular Jacobian matrix at each point, the rank of F is the same at x and $a^{-1}x$, hence constant. By Theorem 5.8, $\ker F = F^{-1}(e_2)$ is a closed regular submanifold whose dimension is $\dim G_1 - \operatorname{rank} F$. From Theorem 6.6, $\ker F$ is a Lie group since it is a regular submanifold (and a group). ∎

(6.15) Example A very useful example of a submanifold which is not regular but is a subgroup of a Lie group is obtained as follows: Let $T^2 = S^1 \times S^1$ and let $F: \boldsymbol{R}^2 \to T^2$ be given by $F(x^1, x^2) = (\exp 2\pi i x^1, \exp 2\pi i x^2)$ as in Example 6.13. Then F is a C^∞ map of rank 2 everywhere and is a homomorphism of Lie groups; the rank may be easily computed at $(0, 0)$ and it is constant by Theorem 6.14.

Now let α be an irrational number and define $G: \boldsymbol{R} \to \boldsymbol{R}^2$ by $G(t) = (t, \alpha t)$. Thus G is obviously an imbedding; its image is the line through the origin of slope α. It follows that $H = F \circ G$ is an immersion of $\boldsymbol{R}$ into T^2 since $DH = DF \cdot DG$ has rank 1 for all $t \in \boldsymbol{R}$. Moreover H is one-to-one since $H(t_1) = H(t_2)$ is equivalent to $\exp 2\pi i t_1 = \exp 2\pi i t_2$ and $\exp 2\pi i \alpha t_1 = \exp 2\pi i \alpha t_2$. However, $\exp 2\pi i u = \exp 2\pi i v$ if and only if $u - v$ is an integer. Clearly $t_1 - t_2$ and $\alpha(t_1 - t_2)$ are both integers only if $t_1 = t_2$. Thus $H: \boldsymbol{R} \to T^2$ is a one-to-one immersion and $H(\boldsymbol{R})$ is an immersed submanifold. However, the interesting fact is that $H(\boldsymbol{R})$ is a dense subset of T^2, so it is about as far from being a regular submanifold as one can imagine: for example, as a subspace it is not locally connected at any point.

We shall prove that $H(\boldsymbol{R})$ is dense in T^2. Since F is continuous and onto, a dense subset D of $\boldsymbol{R}^2$ is mapped to a dense subset of T^2. We will show that $D = F^{-1}(H(\boldsymbol{R}))$ is dense. D consists not only of the line of slope α through the origin but of all lines which can be obtained from it by translation by an integral vector in either direction, that is, any points $(x^1 + m, x^2 + n)$, with m, n integers and with $x^1 = t$, $x^2 = \alpha t$, must also be in D since $F(x^1, x^2) = F(x^1 + m, x^2 + n)$. These lines are all parallel to the given one $H(\boldsymbol{R})$. In fact D consists of the union of all lines $t \to (t + m, \alpha t + n)$, that is, all lines with equation

$$x^2 = \alpha x^1 + (n - \alpha m)$$

for arbitrary integers n, m (Fig. III.11).

Obviously, D is dense on the plane if the y-intercepts $(n - \alpha m)$ form a dense subset of the y-axis. Thus we must show that given α, any real number b, and any $\varepsilon > 0$, there is a pair of integers n, m with $|b - (n - \alpha m)| < \varepsilon$.

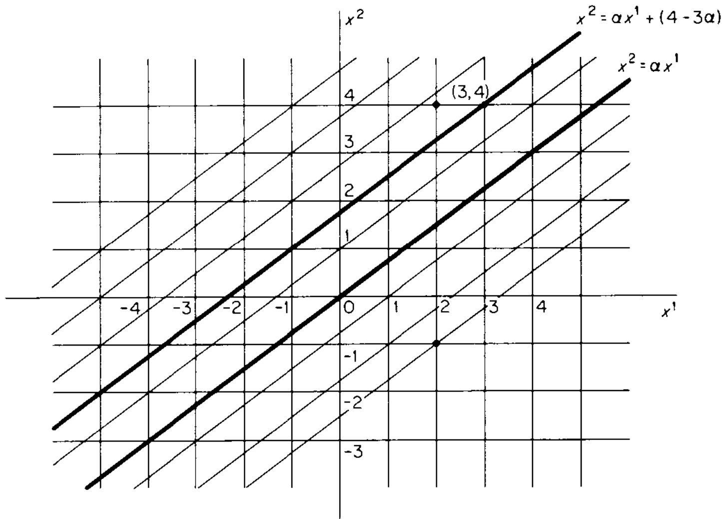


Figure III.11

Assume that there exist integers n', m' such that $0 \leq n' - \alpha m' < \varepsilon$; since $n' - \alpha m'$ is irrational, it must then in fact be positive. It follows that for some integer k, $k(n' - \alpha m') < b \leq (k + 1)(n' - \alpha m')$, which implies $0 < b - k(n' - \alpha m') \leq n' - \alpha m' < \varepsilon$. Since $n - \alpha m = kn' - \alpha km'$ is a y-intercept of a line of D, such that $|b - (n - \alpha m)| < \varepsilon$, the following fact from number theory suffices to complete the proof: Choose n', m' so that $|m'| > 1/\varepsilon$, hence so that $0 < n - \alpha m' < \varepsilon$ or $0 < (-n') - \alpha(-m') < \varepsilon$.

(6.16) *If $\alpha > 0$ is any irrational number, then there exist arbitrarily large integers n', m' such that*

$$\left| \frac{n'}{m'} - \alpha \right| < \frac{1}{m'^2}.$$

A proof is given by Auslander and MacKenzie [1], or Hilbert and Cohn-Vossen [1]. The preceding facts (Example 6.15) are essentially the Kronecker approximation theorem; several beautiful proofs are given by Bohr [1].

We remark that $H: \boldsymbol{R} \to T^2$ in addition to being a one-to-one immersion is a homomorphism of Lie groups so that $\tilde{\boldsymbol{R}} = H(\boldsymbol{R})$ is a subgroup algebraically and an immersed submanifold. It is clearly a Lie group with the manifold structure of $\boldsymbol{R}$. However, it is not a regular submanifold nor is it a closed subset, although it is the image of a closed regular submanifold under a local diffeomorphism F.

(6.17) Definition A (Lie) *subgroup* H of a Lie group G will mean any algebraic subgroup which is a submanifold and is a Lie group with its C^∞ structure as an (immersed) submanifold.

We have already discussed subgroups that are regular submanifolds. We shall prove the following theorem about such subgroups.

(6.18) Theorem *If H is a regular submanifold and subgroup of a Lie group G, then H is closed as a subset of G.*

Proof Let V, ψ be a preferred coordinate neighborhood of e, the identity, i.e., $\psi(e) = (0, \ldots, 0) \in C_\varepsilon^m(0)$ and $g \in H$ if and only if $V \cap H/x^{n+1}(g) = \cdots = x^m(g) = 0$, $n = \dim H$ and $m = \dim G$. If $0 < \delta < \varepsilon$, then $V' = \psi^{-1}(C_\delta^m(0))$, $\bar{V}'$ and $V' \cap H$ are in V. Let $\{h_n\} \subset H$ be a sequence such that $\lim_{n\to\infty} h_n = g$. We must show that $g \in H$, hence $\bar{H} = H$. For $0 < \delta < \varepsilon$ choose W a neighborhood of e sufficiently small so that $W^{-1}W = \{g_1^{-1}g_2 | g_1, g_2 \in G\} \subset V'$ (cf. Exercise 1). Then there exists $N > 0$ such that $L_{g^{-1}}(h_n) = g^{-1}h_n \in W$ for $n > N$, and thus $(g^{-1}h_m)^{-1}g^{-1}h_n = h_m^{-1}h_n \in W^{-1}W \subset V'$. By continuity of the group product, $\lim h_m^{-1}h_n = h_m^{-1}g$, but since $h_m^{-1}h_n \in V'$ for n, $m > N$, we see that $h_m^{-1}h_n g \in \bar{V}' \subset V$. On the other hand $x^j(h_m^{-1}h_n) \equiv 0$ for $j = n+1, \ldots, m$, and hence $x^j(h_m^{-1}g) = 0$ for the same j. Since this point is in V, it follows that $h_m^{-1}g \in H$, which implies $g \in H$. This completes the proof. [Example 4.7 shows that this theorem does not hold for submanifolds in general.] ∎

(6.19) Remark A converse statement is also true: A Lie subgroup H of G that is closed as a subset, i.e., $\bar{H} = H$, is necessarily a regular submanifold (see Lemma IV.9.7). Thus closed Lie subgroups play a special role in Lie group theory (cf. also Theorem 6.14)—although Example (6.15) shows that this property is not preserved by homomorphisms. It is possible, in fact, to prove a much stronger statement: If H is an algebraic subgroup of G and a closed subset, then G is a Lie subgroup—and being closed, a regular submanifold. For a proof of this important theorem (somewhat beyond the scope of this text), see Helgason [1, Chapter II, Section 2], Chevalley [1], or Adams [1, 2.27] (after studying Chapter IV!). Nevertheless, it motivates and validates terminology which we use hereafter: A subgroup H of a Lie group G, which is a regular submanifold, will be called a *closed subgroup* of G.

Exercises

1. Show that given any neighborhood U of e, the identity of a Lie group G, there exists a neighborhood V of e such that $VV^{-1} \subset U$, and a neighborhood W of e such that $W^2 = WW = \{g_1g_2 | g_1, g_2 \in W\} \subset U$.
2. Show that the collection $\{xU\}$, over all neighborhoods U of e, is a base of neighborhoods for x (similarly for $\{Ux\}$).

3. Let A be an arbitrary subset and U an open subset of a Lie group. Show that $AU = \{au \mid a \in A, u \in U\}$ is open.
4. Prove Theorem 6.4 and also that the projections p_1, p_2 and injections i, j of Exercise 3.3, with $(a, b) = (e_1, e_2)$ the identity of $G_1 \times G_2$, are homomorphisms of Lie groups.
5. In Lemma 6.7 show that if N is an immersed submanifold and F is assumed to be continuous as a mapping into N, then F is C^∞ as a mapping into N.
6. Show that if G is a Lie group, $a \in G$, then the map $I_a\colon G \to G$, defined by $I_a(x) = axa^{-1}$, is an automorphism of G.
7. Show that two one-to-one C^∞ immersions $F_1, F_2\colon N' \to M$ with the same image N are equivalent in the sense of Exercise 4.4 if they induce the same topology on N. [*Hint*: Use Exercise 5.]
8. Show that if H is an algebraic subgroup of a Lie group G, then its closure $\bar{H}$ is also an algebraic subgroup. What does Remark 6.19 then imply?
9. In Example 6.15, assume that α (used in the definition of G) is a rational number. Show that $H(\boldsymbol{R})$ is then a regular submanifold of T^2 diffeomorphic to S^1. (In fact, it will be a *knotted* S^1; see Seifert and Threlfall [1, Section 52].

7 The Action of a Lie Group on a Manifold. Transformation Groups

The definition of a group as a set of objects with a law of composition satisfying certain axioms is a relatively recent development. Historically, groups arose as collections of permutations or one-to-one transformations of a set X onto itself with composition of mappings as the group product; for if X is any set whatsoever, then the collection $S(X)$ of all of its "permutations"—in this broad sense—is easily seen to be a group with respect to composition of permutations as product. The same is true for any subcollection G which contains, together with each transformation $\sigma\colon X \to X$, its inverse σ^{-1}, and which contains the composition $\sigma \circ \tau$ of any two of its elements σ and τ. In particular, if X contains just n elements, then $S(X)$ is the symmetric group on n letters and has $n!$ elements, the one-to-one transformations of X onto itself. This was, for example, the point of view of Galois [1], who considered groups of permutations of the roots of a polynomial. Much later Klein [1] discovered the central role of groups in all of the classical geometries: Euclidean, projective, and hyperbolic (non-Euclidean). In this approach, to each geometry is associated a group of transformations or permutations of the underlying space of the geometry, and in each case the geometry with its theorems may be derived from a knowledge of the underlying group. For example, the group G of Euclidean plane geometry is the subgroup of $S(\boldsymbol{E}^2)$ which leaves distances invariant: If $x, y \in \boldsymbol{E}^2$ and

$d(x, y)$ is their distance, then a transformation $T: \boldsymbol{E}^2 \to \boldsymbol{E}^2$ is in the group if and only if $d(Tx, Ty) = d(x, y)$ for all x, y. This is called the *group of rigid motions* and it is generated by translations, rotations, and reflections.

Even though the concept of group does not appear in Euclid's axioms, that of congruence does. These ideas are intimately related: Two figures are congruent if and only if there is a rigid motion σ in G which carries one figure onto the other. It is a consequence of the properties mentioned above for $\sigma, \tau \in G$ that congruence is an equivalence relation.

Although the interpretation of groups as transformation groups of a space has been superseded in algebra, it is still very important in geometry and we shall need to discuss it in various aspects. We first define a very slight generalization of the group $S(X)$.

(7.1) Definition Let G be a group and X a set. Then G is said to *act on* X (on the left) if there is a mapping $\theta: G \times X \to X$ satisfying two conditions:

(i) If e is the identity element of G, then

$$\theta(e, x) = x \qquad \text{for all } x \in X.$$

(ii) If $g_1, g_2 \in G$, then

$$\theta(g_1, \theta(g_2, x)) = \theta(g_1 g_2, x) \qquad \text{for all } x \in X.$$

When G is a topological group, X is a topological space, and θ is continuous, then the action is called *continuous*. When G is a Lie group, X is a C^∞ manifold, and θ is a C^∞, we speak of a C^∞ *action*. C^∞ action is *a fortiori* continuous.

As a matter of notation we shall often write gx for $\theta(g, x)$ so that (ii) reads $(g_1 g_2)x = g_1(g_2 x)$. We also let $\theta_g(x)$ denote the mapping $\theta_g: X \to X$ defined by $\theta_g(x) = \theta(g, x)$, g fixed, so that (ii) may also be written $\theta_{g_1 g_2} = \theta_{g_1} \circ \theta_{g_2}$. When we define *right* action, (i) and (ii) become:

$$\text{(i)} \quad \theta(x, e) = x \qquad \text{and} \qquad \text{(ii)} \quad \theta(\theta(x, g_1), g_2) = \theta(x, g_1 g_2).$$

Usually we are concerned with left action, but in both cases we usually say G *acts* on X, and leave the rest to be determined by the context.

Note that $\theta_{g^{-1}} = (\theta_g)^{-1}$ since $\theta_{g^{-1}} \circ \theta_g = \theta_{g^{-1}g} = \theta_e = i_X$, the identity map on X by (i). This means that each mapping θ_g is one-to-one onto, since it has an inverse. This and (ii) show that the following statement holds:

(7.2) *If G acts on a set X, then the map $g \to \theta_g$ is a homomorphism of G into $S(X)$. Conversely, any such homomorphism determines an action with $\theta(g, x) = \theta_g(x)$.*

We note that the homomorphism is injective if and only if θ_g being the identity implies that $g = e$. If this is so, we shall call the action *effective*. When the action is effective, G may be identified with a subgroup of $S(X)$ by this map $g \to \theta_g$ so that we have precisely the situation discussed in the beginning of the paragraph. Needless to say, these considerations all refer only to the set-theoretic aspects, since $S(X)$ has not been topologized.

We also note that if X is a topological space (C^∞ manifold), G a topological group (Lie group), and the action is continuous (C^∞), then each θ_g is a homeomorphism (diffeomorphism).

(7.3) Example Let H, G be groups, and $\psi: H \to G$, a homomorphism. It is easy to check that $\theta: H \times G \to G$ defined by $\theta(h, x) = \psi(h)x$ is a left action:

(i) $\theta(e, x) = \psi(e)x = x$, since ψ takes the identity of H to the identity of G, and

(ii) $\theta(h_1, \theta(h_2, x)) = \theta(h_1, \psi(h_2)x) = \psi(h_1)(\psi(h_2)x)$ and

$$\theta(h_1 h_2, x) = \psi(h_1 h_2)x = (\psi(h_1)\psi(h_2))x.$$

These agree by the associative law in G. If H and G are Lie groups and ψ is a homomorphism of Lie groups, then the action is C^∞. This may be applied to the case where H is a Lie subgroup of G (or even if $H = G$); in this case ψ is the identity (inclusion) mapping of H into G and we say that H acts on G by left translations.

(7.4) Example A rather simple but important example is known as the *natural action* of $Gl(n, \mathbf{R})$ on $\mathbf{R}^n$: We let $G = Gl(n, \mathbf{R})$ and $X = \mathbf{R}^n$ and we define $\theta: G \times \mathbf{R}^n \to \mathbf{R}^n$ by $\theta(A, x) = Ax$, this being multiplication of the $n \times n$ matrix A by the $n \times 1$ column vector obtained by writing $x \in \mathbf{R}^n$ vertically. This satisfies (i) and (ii) rather trivially, (ii) being again associativity (of matrix products):

$$(AB)x = A(Bx).$$

Since $\theta: G \times \mathbf{R}^n \to \mathbf{R}^n$ is given by polynomials in the entries of $A \in Gl(n, \mathbf{R})$ and $x \in \mathbf{R}^n$, it is a C^∞-map:

$$\theta\left((a_{ij}), \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}\right) = \left(\sum_{j=1}^{n} a_{ij} x^j\right).$$

Now suppose that $H \subset Gl(n, \mathbf{R})$ is a subgroup in the sense of Lie groups, that is, H has its own Lie group structure such that the inclusion map $i: H \to Gl(n, \mathbf{R})$ is an immersion, or—if H is a closed subgroup—an imbedding. Then θ restricted to H defines a C^∞ action $\theta_H: H \times \mathbf{R}^n \to \mathbf{R}^n$. This is

because $\theta_H = \theta \circ i$, $i: H \to G$ is the inclusion map, and both θ and i are C^∞. Using this idea we may give further examples.

(7.5) Example Let $H \subset Gl(2, \boldsymbol{R})$ be the subgroup of all matrices of the form $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$ with $a > 0$. Then H is seen to be a two-dimensional submanifold of $Gl(2, \boldsymbol{R})$ and therefore is a closed subgroup. The restriction to H of the natural action of $Gl(2, \boldsymbol{R})$ on $\boldsymbol{R}^2$ is just

$$\theta_H\left(\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}, \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}\right) = \begin{pmatrix} ax^1 + bx^2 \\ ax^2 \end{pmatrix},$$

which is obviously C^∞, as expected.

(7.6) Example Identify $\boldsymbol{E}^n$ with $\boldsymbol{R}^n$ and let $d(x, y) = (\sum_{i=1}^n (x^i - y^i)^2)^{1/2}$ be the usual metric. The group G of all rigid motions, that is, diffeomorphisms $T: \boldsymbol{R}^n \to \boldsymbol{R}^n$ such that $d(Tx, Ty) = d(x, y)$ is given by transformations T of the form

$$(*) \qquad T(x) = Ax + b,$$

where $A \in O(n)$ and $b \in \boldsymbol{R}^n$—a rotation† A of $\boldsymbol{R}^n$ about the origin followed by a translation taking the origin to b. The group operation is composition of rigid motions.

The group of rigid motions is a Lie group. It is in one-to-one correspondence with $O(n) \times \boldsymbol{R}^n$ and takes its manifold structure from this correspondence, which is given by assigning to each rigid motion $(*)$ the pair $(A, b) \in O(n) \times \boldsymbol{R}^n$. [However, G is not a direct product in the group theoretic sense (Exercise 6).] Since $\theta: G \times \boldsymbol{R}^n \to \boldsymbol{R}^n$ defined by $\theta((A, b), x) = Ax + b$, θ is a C^∞ mapping. Verification that θ defines an action is left to the exercises.

(7.7) Definition Let a group G act on a set M and suppose that $A \subset M$ is a subset. Then GA denotes the set $\{ga \mid g \in G \text{ and } a \in A\}$. The *orbit* of $x \in M$ is the set Gx. If $Gx = x$, then x is a *fixed point* of G; and if $Gx = M$ for some x, then G said to be *transitive* on M. In this case $Gx = M$ for all x.

(7.8) Example Consider the natural action of $Gl(n, \boldsymbol{R})$ on $M = \boldsymbol{R}^n$. The origin 0 is a fixed point of $Gl(n, \boldsymbol{R})$ and $Gl(n, \boldsymbol{R})$ is transitive on $\boldsymbol{R}^n - \{0\}$. For if $x = (x^1, \ldots, x^n) \neq 0$, then there is a basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ with $x = \mathbf{f}_1$. If we express these basis elements in terms of the canonical basis $\mathbf{f}_i = \sum_{j=1}^n a_{ij}\mathbf{e}_j$, $i = 1, \ldots, n$, then we see that $x = A \cdot \mathbf{e}_1$, $A = (a_{ij}) \in Gl(n, \boldsymbol{R})$. From this it follows that every $x \neq 0$ is in the orbit of $\mathbf{e}_1$. This action is not very interesting from the point of view of its orbits. However, if we consider this action

† Including reflections: A is a *proper* rotation if $\det A = +1$.

restricted to various subgroups $G \subset Gl(n, \boldsymbol{R})$, then the orbits can be quite complicated. A relatively simple case of this type is obtained by letting $G = O(n)$, the subgroup of $n \times n$ orthogonal matrices in $Gl(n, \boldsymbol{R})$. This is a closed subgroup as we have seen, and the natural action of $Gl(n, \boldsymbol{R})$ restricted to $O(n)$ is a C^∞ action by Example 7.4. The orbits are the concentric spheres with the origin being a fixed point (sphere of radius zero).

(7.9) Remark The same facts from linear algebra that we used above also show that $Gl(n, \boldsymbol{R})$ is transitive on the collection $\mathbf{B}$ of all bases of $\boldsymbol{R}^n$. Given any basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$, then there exists $A \in Gl(n, \boldsymbol{R})$ such that $A \cdot \mathbf{e}_i = \mathbf{f}_i$, in fact there is exactly one such A. Thus, letting $\mathbf{f} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ and $\mathbf{e} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be elements of $\mathbf{B}$, we may define a left action of $Gl(n, \boldsymbol{R})$ on $\mathbf{B}$, that is, a mapping $\theta\colon Gl(n, \boldsymbol{R}) \times \mathbf{B} \to \mathbf{B}$ by

$$\theta(A, \mathbf{e}) = A \cdot \mathbf{e} = \mathbf{f} = \{A\mathbf{e}_1, \ldots, A\mathbf{e}_n\}.$$

This action is transitive as mentioned, moreover the uniqueness of A (such that $A \cdot \mathbf{e} = \mathbf{f}$) implies that it is *simply transitive*, that is, given bases $\mathbf{f}, \tilde{\mathbf{f}}$, there is exactly one $A \in Gl(n, \boldsymbol{R})$ such that $A \cdot \mathbf{f} = \tilde{\mathbf{f}}$. This means that $Gl(n, \boldsymbol{R})$ is in one-to-one correspondence with $\mathbf{B}$: $A \in Gl(n, \boldsymbol{R})$ corresponds to $A \cdot \mathbf{e}$, where $\mathbf{e}$ is the canonical basis. We may use this correspondence to give $\mathbf{B}$ the topology and C^∞ structure which makes it diffeomorphic to $Gl(n, \boldsymbol{R})$. As a C^∞ manifold it is called the *space of frames* of $\boldsymbol{R}^n$.

We have already mentioned quotient spaces of an equivalence relation as a possible source of manifolds and, in fact, we have produced two examples of such: projective spaces and Grassman manifolds. The most useful and important source of such spaces is furnished by the action of groups on manifolds; at the moment we can only consider the topological aspects, and then only in part, as a preview of things to come.

As a matter of notation we let G denote a Lie group, M a C^∞ manifold, and we assume a C^∞ action $\theta\colon G \times M \to M$. We define a relation $\sim$ on M by $p \sim q$ if for some $g \in G$ we have $q = \theta_g(p) = gp$. It is easily seen that $\sim$ is an equivalence relation and that the equivalence classes coincide with the orbits of G. In fact, $p \sim p$ since $p = ep$ and $p \sim q$ means $q = gp$, which implies $p = g^{-1}q$ or $q \sim p$, so that the relation is reflexive and symmetric. Finally, given that $p \sim q$ and $q \sim p$, we must have $q = gp$ and $r = hq$ so that $r = (hg)p$ and then $p \sim r$. Obviously, $p \sim q$ implies that p and q are on the same orbit, so the equivalence class $[p] \subset Gp$. Conversely, if $q \in Gp$, then $p \sim q$ so $Gp \subset [p]$.

We denote by M/G the set of equivalence classes; it will always be taken with the quotient topology (Definition 2.1) and will often be called the *orbit space* of the action. With this topology the projection $\pi\colon M \to M/G$ (taking each $x \in M$ to its orbit) is continuous, and since the action θ is continuous, π

is also open: If $U \subset M$ is an open set, then so is $\theta_g(U)$ for every $g \in G$ and hence $GU = [U] = \bigcup_{g \in G} \theta_g(U)$, being a union of open sets, is open. The orbit space need not be Hausdorff—but if it is, then the orbits must be closed subsets of M (each is the inverse image by π of a point of M/G and points are closed in a Hausdorff space). We shall be particularly interested in discovering examples in which M/G is a C^∞ manifold and π a C^∞ mapping.

(7.10) Example When $M = \mathbf{R}^n$ and $G = O(n)$ acting naturally as a subgroup of $Gl(n, \mathbf{R})$, then the orbits correspond to concentric spheres and thus are in one-to-one correspondence with the real numbers $r \geq 0$ by the mapping which assigns to each sphere its radius. This is a homeomorphism of $\mathbf{R}^n/O(n)$ and the ray $0 \leq r < \infty$; this is not a manifold, but it is almost one.

(7.11) Example Let G be a Lie group and H a subgroup (in the algebraic sense). Then H acts on G on the right by right translations. If H is a Lie subgroup, then according to Example 7.3 this is a C^∞ action; the set G/H of left cosets coincides with the orbits of this action and is thus a space with the quotient topology. We have the following facts concerning G/H (with this topology).

(7.12) Theorem *The natural map $\pi: G \to G/H$, taking each element of G to its orbit, that is, to its left coset, is not only continuous but open; G/H is Hausdorff if and only if H is closed.*

Proof Since this space—usually called the (left) *coset space*—coincides with the orbit space of H acting on G, π is continuous and open. To prove the last statement we use the C^∞ mapping $F: G \times G \to G$ defined by $F(x, y) = y^{-1}x$. Since F is continuous and $F^{-1}(H)$ is the subset $R = \{(x, y) \mid x \sim y\}$ of $G \times G$, we see by Lemma 2.4 that R is closed and G/H is Hausdorff if and only if H is a closed subset of G. ∎

We conclude this section with two definitions, using terminology which we justify in the exercises.

(7.13) Definition Let G be a group acting on a set X and let $x \in X$. The *stability* or *isotropy* group of x, denoted by G_x, is the subgroup of all elements of G leaving x fixed, $G_x = \{g \in G \mid gx = x\}$.

(7.14) Definition Let G, X be as in the previous definition. Then G is said to act *freely* on X if $gx = x$ implies $g = e$, the identity, that is, the identity is the only element of G having a fixed point.

Exercises

1. Show that if G acts on X as in Definition 7.13, then for each x, G_x is a subgroup of G, which in the case of continuous action is a closed subset of G.
2. Suppose that G acts transitively on X. Then given $x, y \in X$, prove that G_x and G_y are conjugate subgroups of G.
3. Let G act transitively on X and let x_0 be a point of X. Define $\tilde{F}: G \to X$ by $\tilde{F}(g) = gx_0$. Prove: (i) that there is a unique one-to-one mapping $F: G/G_{x_0} \to X$ such that $\tilde{F} = F \circ \pi$, $\pi: G \to G/G_{x_0}$, the natural projection to cosets, (ii) that $\tilde{F}$ and F are continuous if the action is continuous and in this case F is a homeomorphism if and only if $\tilde{F}$ is open.
4. Show that $O(n)$ acts transitively on S^{n-1}, the unit sphere of $\boldsymbol{R}^n$, in a natural way and determine the isotropy subgroup.
5. Show that $Gl(n, \boldsymbol{R})$ acts transitively on $P^{n-1}(\boldsymbol{R})$ and determine the isotropy subgroup of $[(1, 0, \ldots, 0)]$.
6. Let $G = O(n) \times V^n$ and define a product in G by $(A, v)(B, w) = (AB, Bv + w)$. Prove that G is a Lie group and acts on $\boldsymbol{R}^n$ by $(A, v) \cdot x = Ax + v$. Show that $\{I\} \times V^n$, I being the identity matrix, is a closed submanifold and normal subgroup and that $O(n) \times \{0\}$ is a closed submanifold and subgroup of G.
7. Let G be the set of 2×2 matrices of the form $\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)$, where $a > 0$ and b are real numbers. Show that G is a Lie group and acts on $\boldsymbol{R}$ by

$$\theta\left(\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}, x\right) = ax + b.$$

8. Let H, a subgroup of G, act on G by left translations. Prove that this is a free action. Show that if G acts freely and transitively on the left on X, then G and X are in one-to-one correspondence and if they are identified, the action is equivalent to left translations.
9. Let the multiplicative group of nonzero real numbers $\boldsymbol{R}^*$ act on $\boldsymbol{R}_0^{n+1} = \boldsymbol{R}^{n+1} - \{0\}$ by $\theta: \boldsymbol{R}^* \times \boldsymbol{R}_0^{n+1} \to \boldsymbol{R}_0^{n+1}$, defined by $\theta(t, x) = tx$. Show that $\boldsymbol{R}_0^{n+1}/\boldsymbol{R}^*$ is homeomorphic to $P^n(\boldsymbol{R})$.
10. Let G be a Lie group and H be a closed subgroup and define a left action $\theta: G \times G/H \to G/H$ by $\theta(g, xH) = (gx)H$. Show that this action of G on the coset space G/H is continuous and that the isotropy group of $[e] = H$ is exactly H itself.
11. As in Example 2.6 identify the k-frames in $\boldsymbol{R}^n$, $F(k, n)$, with $k \times n$ matrices of rank k. Show that $Gl(k, R)$ acts on $F(k, n)$ by left multiplication with $G(k, n)$ as orbit space and that the right action of $Gl(n, \boldsymbol{R})$ on $F(k, n)$ induces a transitive action on $G(k, n)$.

8 The Action of a Discrete Group on a Manifold

We will consider in some detail what might seem to be the simplest case in which we could hope to use group action to obtain new examples of manifolds via the quotient or orbit space concept discussed in the previous section. By a *discrete* group Γ we shall mean a group with a countable number of elements and the discrete topology (every point is an open set). The countability means that Γ falls within our definition of a manifold: it has a countable basis of open sets each homeomorphic to a zero-dimensional Euclidean space, that is, a point. Thus Γ is a zero-dimensional Lie group. In this case to verify that an action $\theta: \Gamma \times \tilde{M} \to \tilde{M}$ is C^∞, we need only show that for each $h \in \Gamma$ the mapping $\theta_h: \tilde{M} \to \tilde{M}$ is a diffeomorphism. For convenience of notation, we will let h denote θ_h, writing hx for $\theta_h(x)$, and so on. We suppose then that a C^∞ action is given and consider the set of orbits $M = \tilde{M}/\Gamma$ with the quotient topology discussed before: $U \subset M$ is open if and only if $\pi^{-1}(U)$ is open in $\tilde{M}$, where $\pi: \tilde{M} \to M$ denotes the natural map taking each x to its orbit Γx; we have seen that π is then continuous and open.

If M is Hausdorff in this topology, then points are closed sets and the inverse image of any $p \in M$, that is, the orbit $\pi^{-1}(p)$, must be closed. Thus an obvious necessary condition for M to possess some kind of reasonable topology and manifold structure is that for each $x \in \tilde{M}$ the orbit Γx is closed. However, this condition is not sufficient. A stronger requirement is the following: *Given any point $x \in \tilde{M}$ and any sequence $\{h_n\}$ of distinct elements of Γ, then $\{h_n x\}$ does not converge to any point of $\tilde{M}$.* A group action with this property is called *discontinuous*; it is equivalent to the requirement that each orbit be a closed, discrete subset of $\tilde{M}$. In the presence of other conditions this is sometimes enough to ensure that $\tilde{M}/\Gamma$ is Hausdorff (see Exercise 2), but in general we need the following condition, which is even stronger:

(8.1) Definition A discrete group Γ is said to act *properly discontinuously* on a manifold $\tilde{M}$ if the action is C^∞ and satisfies the following two conditions:

(i) Each $x \in \tilde{M}$ has a neighborhood U such that the set $\{h \in \Gamma \mid hU \cap U \neq \varnothing\}$ is finite;

(ii) If $x, y \in \tilde{M}$ are not in the same orbit, then there are neighborhoods U, V of x, y such that $U \cap \Gamma V = \varnothing$.

Observe that (ii) implies at once that $M = \tilde{M}/\Gamma$ is Hausdorff: In fact it is equivalent to the statement that the subset $R = \{(x, y) \mid x \sim y\} \subset M \times M$ is closed (compare Lemma 2.4).

A consequence of proper discontinuity is the following statement whose proof is left as an exercise (Exercise 3)—it could be used to replace (i) in the definition, so we denote it by (i′).

(i′) *The isotropy group Γ_x of each $x \in \tilde{M}$ is finite, and each x has a neighborhood U such that $hU \cap U = \varnothing$ if $h \notin \Gamma_x$ and $hU = U$ if $h \in \Gamma_x$.*

(8.2) Example Let $M = S^{n-1}$, the set $\{x \in \boldsymbol{R}^n \mid \|x\| = 1\}$ and $\Gamma = \boldsymbol{Z}_2$, the cyclic group of order 2 with generator h, that is, Γ consists of h and $h^2 = e$, the identity. Then $h(x) = -x$ and $e(x) = x$ defines an action of Γ on S^{n-1}. It is left as an exercise to see that the action $\theta: \boldsymbol{Z}_2 \times S^{n-1} \to S^{n-1}$ is free and properly discontinuous and that the quotient space $S^{n-1}/\boldsymbol{Z}_2$ is none other than real projective $n-1$ space $P^{n-1}(\boldsymbol{R})$.

This and other examples, such as S^1 identified with $\boldsymbol{R}/\boldsymbol{Z}$ (see Example 8.7 below), lead to the following theorem:

(8.3) Theorem *Let Γ be a discrete group which acts freely and properly discontinuously on a manifold $\tilde{M}$. Then there is a unique C^∞ structure of differentiable manifold on $M = \tilde{M}/\Gamma$ (with the quotient topology) such that each $p \in M$ has a connected neighborhood U with the property: $\pi^{-1}(U) = \bigcup \tilde{U}_\alpha$ is a decomposition of $\pi^{-1}(U)$ into its (open) connected components and $\pi \mid \tilde{U}_\alpha$ is a diffeomorphism onto U for each component $\tilde{U}_\alpha$.*

Proof The manifold M is Hausdorff since Γ acts properly discontinuously. By Lemma 2.3 it has a countable basis of open sets. Using both (i′) and the assumption that the action is free, we may find for each $x \in \tilde{M}$ a neighborhood $\tilde{U}$ such that $h\tilde{U} \cap \tilde{U} = \varnothing$ except when $h = e$. This implies that $\pi_{\tilde{U}}$ $(= \pi \mid \tilde{U})$ is one-to-one onto its image U, and therefore $\pi_{\tilde{U}}: \tilde{U} \to U$ is a homeomorphism of $\tilde{U}$ to the open set U—the mapping π being both continuous and open. There is no loss of generality in supposing $\tilde{U}$ to be a connected coordinate neighborhood $\tilde{U}, \tilde{\varphi}$. Then taking $\varphi = \tilde{\varphi} \circ \pi_{\tilde{U}}^{-1}$, we have $\varphi: U \to \tilde{\varphi}(\tilde{U}) \subset \boldsymbol{R}^n$ is a homeomorphism. Since every $p \in M$ is the image of some $x \in \tilde{M}$, we see that M is locally Euclidean. Thus M is a topological manifold. The coordinate neighborhoods U, φ just described will be called *admissible*: If they are C^∞ compatible, then they determine a C^∞ structure on M for which π is C^∞, in fact is a local diffeomorphism. Suppose that $U = \pi(\tilde{U})$ and $V = \pi(\tilde{V})$ are overlapping admissible neighborhoods with coordinate maps $\varphi = \tilde{\varphi} \circ \pi_{\tilde{U}}^{-1}$ and $\psi = \tilde{\psi} \circ \pi_{\tilde{V}}^{-1}$. Although $\tilde{U} \cap \tilde{V}$ may be empty, there must be an $h \in \Gamma$ such that $\tilde{U} \cap h\tilde{V} \neq \varnothing$. Since $\pi = \pi \circ h$, we may write $\psi^{-1} = \pi \circ \tilde{\psi}^{-1} = \pi \circ h \circ \tilde{\psi}^{-1}$; hence $\varphi \circ \psi^{-1} = \tilde{\varphi} \circ \pi_{\tilde{U}}^{-1} \circ \pi \circ h \circ \tilde{\psi}^{-1} = \tilde{\varphi} \circ h \circ \tilde{\psi}^{-1}$. These are all C^∞ maps, in fact diffeomorphisms, and hence $\varphi \circ \psi^{-1}$ is C^∞. The same holds for $\psi \circ \varphi^{-1}$, which proves that the admissible neighborhoods U, φ define a C^∞ structure. The coordinates of $p \in \tilde{U}$ and $F(p) \in U$ being exactly the same, $\pi | \tilde{U}$ is a diffeomorphism, and hence $\pi = \pi \circ h$ is a diffeomorphism of $h\tilde{U}$ onto U for any $h \in \Gamma$. Since $\{\pi^{-1}(U)\}$ covers $\tilde{M}$, $\pi: \tilde{M} \to M$ is a C^∞ map and a local diffeomorphism. Because of the requirement that $\pi(\tilde{U})$ be a diffeomorphism, no other C^∞ structure is possible. ∎

We remark that π is C^∞ of rank $n = \dim \tilde{M} = \dim M$ since it is locally a diffeomorphism. Of course, it can be one-to-one only when $\Gamma = \{e\}$ for $\pi^{-1}(p) = \Gamma x$ for some x, and this orbit Γx is in one-to-one correspondence with Γ itself by virtue of the assumption that Γ acts freely.

We shall prove a lemma and then a theorem which will supply some examples of free, properly discontinuous action.

(8.4) Lemma *Let G be a Lie group and Γ an algebraic subgroup. Then there is a neighborhood U of the identity e such that $\Gamma \cap U = \{e\}$ if and only if Γ is a discrete subspace, in which case $\bar{\Gamma} = \Gamma$, i.e., Γ is closed and countable.*

Proof Clearly, if Γ is a discrete subspace, then every $h \in \Gamma$ has a neighborhood V in G such that $V \cap \Gamma = \{h\}$. Conversely, if $e \in \Gamma$ has such a neighborhood, say U, then for any $h \in \Gamma$, $V = hU$ contains only the element h of Γ. It is even possible to choose then a neighborhood W of e such that $WW^{-1} \subset U$, in which case $h_1 W \cap h_2 W = \varnothing$ unless $h_1 = h_2$. For if $h_1 w_1 = h_2 w_2$ is the intersection, then $h_2^{-1} h_1 = w_2 w_1^{-1} \in W \subset U$ which implies $h_2^{-1} h_1 = e$ or $h_1 = h_2$. Then Γ must be countable since $\{hW \mid h \in \Gamma\}$ is a nonintersecting family of open sets in one-to-one correspondence with Γ. Any such family is countable since G satisfies the second axiom of countability. Finally, if Γ is not a closed subset, it is easy to see that there must be a sequence $\{h_n\} \subset \Gamma$ with $\lim_{n\to\infty} h_n = e$. This contradicts the existence of U with $U \cap \Gamma = \{e\}$. ∎

We remark that a Γ with this property is a closed zero-dimensional Lie subgroup of G in the sense of Definition 6.17; such subgroups are often called simply *discrete subgroups*. We give examples below.

(8.5) Theorem *Any discrete subgroup Γ of a Lie group G acts freely and properly discontinuously on G by left translations.*

Proof No other translation than the identity has a fixed point so the action is free. To see that it is properly discontinuous we must check (i) and (ii) of Definition 8.1. Choosing U, V neighborhoods of e as in the proof of the preceding lemma so that $VV^{-1} \subset U$ and $U \cap \Gamma = \{e\}$, we see that the only $h \in \Gamma$ such that $hV \cap V \neq \varnothing$ is $h = e$. This proves (i). To prove (ii) we argue as follows. If Γx and Γy are distinct orbits, then $x \notin \Gamma y$, and since Γy is closed, by the regularity of G there is a neighborhood U of x such that $U \cap \Gamma y = \varnothing$. Let V be a neighborhood of e such that $xVV^{-1} \subset U$. If the open sets $\Gamma x V$ and $\Gamma y V$ intersect, then some element of xVV^{-1} must be in Γy, which is an immediate contradiction. This completes the proof. ∎

(8.6) Corollary *If Γ is a discrete subgroup of a Lie group G, then the space of right (or left) cosets G/Γ is a C^∞ manifold and $\pi\colon G \to G/\Gamma$ is a C^∞ mapping.*

This is a combination of Theorems 8.3 and 8.5. It may also be considered as a generalization of Theorem 7.12 since Γ is a closed subgroup of G.

(8.7) Example A particularly important example is the following. Let $G = \boldsymbol{V}^n$, that is, $\boldsymbol{R}^n$ considered as a vector space, and let $\Gamma = \boldsymbol{Z}^n$, the n-tuples of integers—usually called the *integral lattice*. [More generally one could take for Γ the integral linear combinations of any basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ of V^n.] Γ is a discrete subgroup; the neighborhood $C_\varepsilon^n(0)$ of the origin with $\varepsilon < 1$ does not contain any element of Γ other than $(0, \ldots, 0)$. The reader should verify that $\boldsymbol{V}^n/\Gamma = \boldsymbol{V}^n/\boldsymbol{Z}^n$ is diffeomorphic to $T^n = S^1 \times \cdots \times S^1$, the n-dimensional torus, and that π is a Lie group homomorphism of $\boldsymbol{V}^n$ onto T^n with Γ as kernel.

(8.8) Example Any finite subgroup Γ of a Lie group G is a discrete subgroup. When C is compact, a discrete subgroup must be finite; but even in this case there are interesting examples. Thus in the case of $SO(3)$, the group of 3×3 orthogonal matrices of determinant $+1$, thc subgroups of symmetries of the five regular solids give examples among which is the famous icosahedral group, which contains 60 elements. (See Wolf [1, Section 2.6].)

(8.9) Example In the case of groups which are not compact we have many variations of the following theme: Let $G_0 = Gl(n, \boldsymbol{R})$ and $\Gamma_0 = Sl(n, \boldsymbol{Z})$, the $n \times n$ matrices with integer coefficients and determinant $+1$. Since the topology of G_0 is obtained by considering it as an open subset of $\boldsymbol{R}^{n^2}$, it is clear that Γ_0 corresponds to the intersection of G_0 with the integral lattice $\boldsymbol{Z}^{n^2}$ and hence is discrete. Having said this, suppose G to be a Lie subgroup of G_0 and let $\Gamma = \Gamma_0 \cap G$. Then Γ is discrete in G. An illustration is the following: Let G be all matrices in $Gl(n, \boldsymbol{R})$ with $+1$ on the main diagonal and zero below and let Γ be its intersection with $Sl(n, \boldsymbol{Z})$.

An interesting question about which one can speculate is the following: In which, if any, of these cases is G/Γ compact? Note that it is compact when $G = \boldsymbol{V}^n$ and $\Gamma = \boldsymbol{Z}^n$. A necessary and sufficient condition for compactness is the existence of a compact subset $K \subset G$ whose Γ-orbit covers G, $\Gamma K = G$. In Example 8.7, any cube K of side one or greater has this property.

We terminate by mentioning some examples related to "tiling" the plane and to crystallography. Note that reflection in a line is a rigid motion of the plane, and in fact any rigid motion is a product of reflections, which thus generate the group of motions of the plane. For example, the group Γ generated by reflections in the four lines $x = 0$, $x = \frac{1}{2}$, $y = 0$, $y = \frac{1}{2}$ relative to a fixed Cartesian coordinate system contains the group of translations $(x, y) \to (x + m, y + n)$, m, n integers. This latter group may be identified

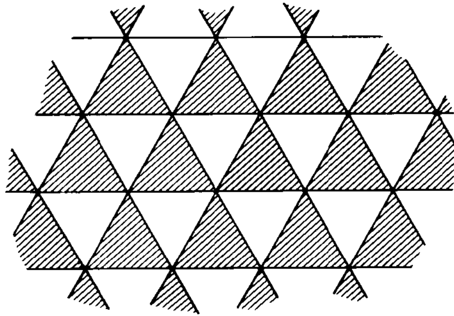


Figure III.12

with the subgroup Z^2 of V^2 discussed above. Moreover the action of Γ leaves unchanged the figure consisting of lines $x = k/2$, $y = l/2$, k, l integers, that is, a collection of squares which "tile" the plane.

Similarly, if we tile the plane with other polygons as in Fig. III.12, we see that the group Γ of reflections in all lines forming edges of these polygons leaves the whole configuration or tiling unchanged. The reader can verify geometrically that the group Γ in these illustrations acts properly discontinuously. Is the action free? This is an important method of obtaining such group actions. It has an immediate extension to E^3, where it is intimately related to crystal structures. (See, e.g., Grove and Benson [1].)

Exercises

1. Let Γ be an abstract group acting as diffeomorphisms on a C^∞ manifold M and assume that Γ_p is finite for every $p \in M$. If the orbits are closed, discrete subsets of M, then prove that Γ is necessarily countable and the action discontinuous.
2. Suppose that a C^∞ manifold M is a metric space and that Γ is a discrete group of C^∞ isometries acting discontinuously on M. Show that the action is necessarily properly discontinuous.
3. (a) Show that (i) may be replaced by (i′) in Definition 8.1; and
   (b) show that (ii) may be replaced by (ii′): $\tilde{M}/\Gamma$ is Hausdorff.
4. Show that the only closed nontrivial subgroups of $\boldsymbol{R}$ are of the form $\Gamma = \{na \mid a \neq 0, n \in Z\}$. Determine all closed subgroups of $\boldsymbol{R}^n$.
5. Verify the statements of Example 8.2.
6. Let G consist of all 3×3 matrices which have $+1$ along the diagonal and zero below and Γ the matrices in G with integer entries. Show that Γ is a closed discrete subgroup and G/Γ is a compact Hausdorff space.
7. Prove that a discrete normal subgroup of a connected Lie group G is in the center of G.

9 Covering Manifolds

Some of the examples of the previous section are intimately related to the notion of covering manifolds. Let $\tilde{M}$ and M be two C^∞ manifolds of the same dimension and $\pi\colon \tilde{M} \to M$ a C^∞ mapping. Using this notation, we make the following definition:

(9.1) Definition $\tilde{M}$ is said to be a *covering* (*manifold*) of M with covering mapping π if π is onto, $\tilde{M}$ is connected and if each $p \in M$ has a connected neighborhood U such that $\pi^{-1}(U) = \bigcup U_\alpha$, a union of open components U_α, with the property that π_{U_α}, the restriction of π to U_α, is a diffeomorphism onto U. The U are called *admissible* neighborhoods and π is called the *projection* or *covering* mapping.

Examples abound in the previous section: $\tilde{M} = \boldsymbol{R}$ covers $M = S^1$ realized as complex numbers of absolute value $+1$ with $\pi(t) = \exp 2\pi i t$. (This may be visualized as in Fig. III.13 with π the projection to the circle S^1.) More generally $\tilde{M} = \boldsymbol{R}^n$ covers T^n. Example 8.2 shows that S^{n-1} covers $P^{n-1}(\boldsymbol{R})$ and in a very general way Theorem 8.3 tells us that if Γ acts freely and properly discontinuously on $\tilde{M}$, then $\tilde{M}$ covers $M = \tilde{M}/\Gamma$. Here the map π is the obvious one: it takes each $x \in \tilde{M}$ to its orbit Γx which is a point of M.

It would seem at first glance that the examples of covering manifolds must be much more extensive than those furnished via Theorem 8.3. We shall see much later that this is not the case. At present we can only begin to show how this is demonstrated. Let us assume then that $\pi\colon \tilde{M} \to M$ is any covering of a manifold M by a connected manifold $\tilde{M}$. We indicate how this may give rise to a group $\tilde{\Gamma}$ acting freely and properly discontinuously on $\tilde{M}$.

(9.2) Definition A diffeomorphism $h\colon \tilde{M} \to \tilde{M}$ is said to be a *covering transformation*, or *deck transformation*, if $\pi \circ h = \pi$.

Note that this is equivalent to the requirement that each set $\pi^{-1}(p)$ is carried into itself. In case the covering is one arising from free, properly discontinuous action of a group Γ on $\tilde{M}$, then each $h \in \Gamma$ is a covering transformation of the covering $\pi\colon \tilde{M} \to \tilde{M}/\Gamma$. More generally, the set $\tilde{\Gamma}$ of all covering transformations acts as a group of diffeomorphisms on $\tilde{M}$. It contains at least the identity so it is not empty.

Given any $x \in \tilde{M}$ and $p = \pi(x)$, let U be an admissible neighborhood of p so $\pi^{-1}(U) = \bigcup \tilde{U}_\alpha$, where $\alpha = 1, 2, \ldots$ (the collection of mutually disjoint neighborhoods $\{\tilde{U}_\alpha\}$ must be countable). Since the U_α are connected components of $\pi^{-1}(U)$, either $h(U_\alpha) \cap U_{\alpha'} = \phi$, or $h(U_\alpha) = U_{\alpha'}$ and $h\colon \tilde{U}_\alpha \to \tilde{U}_{\alpha'}$ is a diffeomorphism—with $h|\tilde{U}_\alpha = \pi_{\tilde{U}_{\alpha'}}^{-1} \circ \pi_{\tilde{U}_\alpha}$. We can conclude that the points

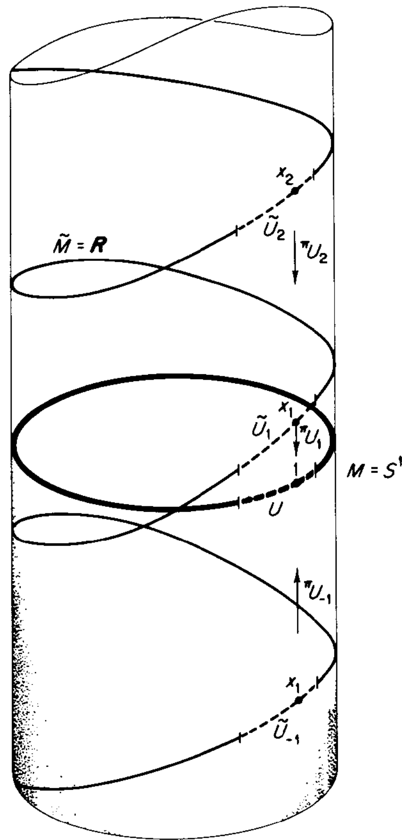


Figure III.13

left fixed by h form an open set. By continuity of h they also form a closed set, and—$\tilde{M}$ being connected—this set is empty or h is the identity. In particular, two covering transformations with the same value on a point x must be identical. Thus covering transformations are completely determined by the permutation $\alpha \to \alpha'$ they induce on the set of points $\{x_\alpha\} = \pi^{-1}(p)$ for an arbitrary (but fixed) point $p \in M$. In particular, the action of $\tilde{\Gamma}$ on $\tilde{M}$ is free. If $x_1 \in \pi^{-1}(p)$, then $h \to hx_1$ maps $\tilde{\Gamma}$ into $\pi^{-1}(p)$. This mapping is an injection so $\tilde{\Gamma}$ must be countable, and as a discrete group of diffeomorphisms of $\tilde{M}$, it acts differentiably on $\tilde{M}$. This proves, in part, the following theorem:

(9.3) Theorem *With the notation above, $\tilde{\Gamma}$ acts freely and properly discontinuously on $\tilde{M}$ and the natural map $\pi_1: \tilde{M} \to \tilde{M}/\tilde{\Gamma}$ is a covering mapping. If $\tilde{\Gamma}$ is a transitive on $\pi^{-1}(p)$ for every $p \in M$, then there is a natural identification of M with $\tilde{M}/\tilde{\Gamma}$ and of π_1 with π.*

Proof We have already seen that $\tilde{\Gamma}$ acts on $\tilde{M}$ freely since only the identity has a fixed point. We must check (using admissible neighborhoods) that the action is properly discontinuous. If $x \in \tilde{M}$ and $p = \pi(x)$, then $x \in \{x_\alpha\} = \pi^{-1}(p)$, say $x = x_1$, and if $h \neq e$, then $h(x_1) = x_\beta \neq x_1$ so $h(\tilde{U}_1) = \tilde{U}_\beta$ with $\tilde{U}_\beta \cap \tilde{U}_1 = \varnothing$. Thus part (i) of proper discontinuity is proved.

For part (ii) we take $x, y \in \tilde{M}$ not in the same orbit of $\tilde{\Gamma}$ and consider two cases: either $\pi(x) = \pi(y)$ or not. If they are the same, denote this point by p, as above, and note that in permuting $\{x_\alpha\} = \pi^{-1}(p)$, no $h \in \tilde{\Gamma}$ takes $x = x_\alpha$ to $y = x_\beta$ $(\alpha \neq \beta)$; whence $\tilde{U}_\alpha$ is not carried to $\tilde{U}_\beta$ by any $h \in \tilde{\Gamma}$. This establishes (ii) of Definition 8.1 in this case. When, on the other hand, $\pi(x) = p$ and $\pi(y) = q$ are distinct, it is even easier. Let U, V be disjoint admissible neighborhoods of p, q, respectively. Then the open sets $\pi^{-1}(U)$ and $\pi^{-1}(V)$ are disjoint and carried into themselves by every $h \in \tilde{\Gamma}$, so they answer to requirement (ii). Thus the action is properly discontinuous and the first part of the theorem follows from Theorem 8.3.

To prove the second part of the theorem we define $\pi_2 : \tilde{M}/\tilde{\Gamma} \to M$ by $\pi_2([x]) = \pi(x)$, where $x \in \tilde{M}$ and $[x]$ is its Γ-orbit. Since $\pi(x) = \pi(h(x))$ for any $h \in \tilde{\Gamma}$, π_2 is well-defined and $\pi = \pi_2 \circ \pi_1$. We leave it as an exercise to verify that $\pi_2 : \tilde{M}/\tilde{\Gamma} \to M$ is a covering. If $\tilde{\Gamma}$ is transitive on $\pi^{-1}(p)$ for each $p \in M$, then $\pi^{-1}(p)$ is a $\tilde{\Gamma}$-orbit and $\pi_2^{-1}(p)$ is a single point of $\tilde{M}/\tilde{\Gamma}$. In this case the latter part of the theorem is a consequence of the following easy lemma applied to π_2. ∎

(9.4) Lemma *Let $\pi : \tilde{M} \to M$ be a covering and suppose that for some $p \in M$, $\pi^{-1}(p)$ is a single point. Then π is a diffeomorphism.*

Exercises

1. Prove Lemma 9.4 by using the connectedness of $\tilde{M}$.
2. Show that $\pi_2 : \tilde{M}/\tilde{\Gamma} \to M$ as defined above is a covering.
3. Show that the covering transformations form a group and that if $x, y \in \tilde{M}$, a covering manifold of M, then there is at most one covering transformation taking x to y. Show further that if $\tilde{\Gamma}$ is transitive on $\pi^{-1}(p)$ for some $p \in M$, then it is transitive for every p.
4. Let $\pi : \tilde{M} \to M$ be a covering and X a connected space $F : X \to M$ a continuous mapping. Suppose $\tilde{F}_1, \tilde{F}_2 : X \to \tilde{M}$ are continuous, have the property that $\pi \circ \tilde{F}_i = F$, and agree on one point of X. Show that $\tilde{F}_1 = \tilde{F}_2$.
5. Let $\pi : \tilde{M} \to M$ be a covering and $F : [a, b] \to M$ a continuous curve from $F(a) = p$ to $F(b) = q$. If $x_0 \in \pi^{-1}(p)$, show that there is a unique continuous curve $\tilde{F} : [a, b] \to \tilde{M}$ such that $\tilde{F}(a) = x_0$ and $\pi \circ \tilde{F} = F$.

6. Let $I = [0, 1]$ be the closed unit interval and $I^n = I \times \cdots \times I$, the n-fold Cartesian product. Suppose $F: I^n \to M$ is continuous with $p = F(0, \ldots, 0)$. If $\pi: \tilde{M} \to M$ is a covering and $x \in \pi^{-1}(p)$, then prove that there is a *unique* continuous map $\tilde{F}: I^n \to \tilde{M}$ such that $F = \pi \circ \tilde{F}$ and $\tilde{F}(0, \ldots, 0) = x$.
7. Let $\tilde{M}, M$ be C^∞ manifolds of dimension n and $\pi: \tilde{M} \to M$ a C^∞ map which is onto and has rank n at each point. Prove or disprove the statements: (a) π is locally a diffeomorphism; (b) π is a covering map.

Notes

The concept of differentiable manifold which was presented in this chapter is the result of many influences and the work of many great mathematicians beginning with Gauss and Riemann. In its present form it is of fairly recent creation. The early work in differential geometry was of a local character—hence open subsets of Euclidean space were adequate models. Even the space of non-Euclidean geometry, as a manifold, is equivalent to Euclidean space—it is only the metric aspects of the geometry which are different. The same is true of Lie groups, as studied by Lie, since only group germs or local Lie groups, that is, neighborhoods of the identity, were considered. Except for Riemann surfaces and projective spaces, there was little to force the global aspects of manifolds into prominence. Nevertheless it seems clear that Riemann, and certainly Poincaré, conceived of manifolds in essentially the same way that we do today although without the present precise separation between the underlying topological space and the differentiable structure. In fact the concept was so difficult that even many great mathematicians —well into this century—were unable to grasp the concept of a topological or differentiable manifold. This depended on the idea of overlapping coordinate neighborhoods, perhaps first clearly expressed in the book on Riemann surfaces by Weyl [1] published in 1923 and brought to its present form in the seminal paper of H. Whitney [1] published in 1936. This paper was particularly important in that it showed that there was no essential distinction between an abstract manifold (Def. 1.2) and a regular submanifold of Euclidean space (cf. Theorem V.4.7).

Poincaré, whose imprint is everywhere in this subject, studied manifolds from many points of view: as phase spaces of dynamical systems, as Riemann surfaces (in which covering spaces and discontinuous groups played an important role), and from the aspect of algebraic topology. (For discussion of these topics the reader can consult the survey article by Smale [1] and the books of Siegel [1] or Lehner [1].) All of the work of Cartan on Lie groups and differential geometry (see Chern and Chevalley [1]) has had an enormous influence on the subject as has the enormous development of algebraic topology, beginning especially with the work of L. E. J. Brouwer and H. Hopf. The latter was particularly interested in the interplay between topological invariants and differential geometric properties, beginning with his paper (Hopf [1]) in 1925.

The work of the past 50 to 60 years is far too extensive to summarize here, however one recent development is of special relevance to this chapter. One basic question left over from the foundations of the subject was a better understanding of the relation between the underlying topological structure of a manifold and its differentiable structure. It was long believed that, up to diffeomorphism, only one differentiable structure could be defined on a topological manifold. In a famous paper in 1956 Milnor [4] showed that this was not in fact the case and that S^7 was a counterexample. Soon after Kervaire and Milnor [1] showed that many spheres fail to have a unique C^∞ structure (see their tabulation below). Then in 1961 Kervaire found an example of topological manifold of dimension 10 with no differentiable structure at all! Finally, very recently

it has been shown that every Euclidean manifold $\boldsymbol{R}^n$, with the exception of $\boldsymbol{R}^4$, has a unique differentiable structure but that the exception $\boldsymbol{R}^4$ has many (see Freedman [1] and Gompf [1]).

A very readable introduction to some of these questions may be found in Milnor [5], a book which also contains a discussion of the Poincaré conjecture by R. H. Bing.

THE DIMENSION n AND NUMBER m OF NONDIFFEOMORPHIC STRUCTURES ON S^n

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
m	1	1	?	1	1	1	28	2	8	6	992	1	3	2	16256	2

IV VECTOR FIELDS ON A MANIFOLD

In this chapter we introduce some of the most basic tools used in the study of differentiable manifolds. First we define the tangent space $T_p(M)$ attached to each $p \in M$, M a C^∞ manifold. Each element X_p of $T_p(M)$ can be considered as an operator (directional derivative) on C^∞-functions at p, generalizing one of the definitions in the case of $\boldsymbol{R}^n$. We also see that a C^∞ mapping $F\colon M \to N$ induces a linear map $F_*\colon T_p(M) \to T_{F(p)}(N)$ on the tangent space at each point. Associated to a coordinate system around p will be a basis of $T_p(M)$. If $\hat{F}$ is the expression for F in local coordinates around p and $F(p)$, then the Jacobian $D\hat{F}$ is the matrix of F_* with respect to these bases.

Assigning a vector X_p to each $p \in M$ we obtain a vector field on M, just as in the special case $\boldsymbol{R}^n$. Vector fields are intimately associated with the action of the Lie group $\boldsymbol{R}$ on M, that is, one-parameter transformation groups. The relation between them is a consequence of—and in some sense equivalent to—the fundamental existence theorem for solutions of systems of ordinary differential equations. Section 2 gives the basic definitions of vector fields, Section 3 the basic definitions of one-parameter transformation groups acting on a manifold and Section 4 the existence theorem. Systems of ordinary differential equations are shown to coincide with vector fields on manifolds and their solutions with curves on the manifold which are tangent to the vectors of the field. These curves are also orbits of the group action. In Section 5 a number of examples are given, using Lie group action as a starting point.

In Section 6 a study is made of one-parameter (one-dimensional) subgroups of a Lie group G. These are basic in the study of Lie groups, but we use them primarily as a source of examples and to illustrate the basic ideas above. They are in one-to-one correspondence with the vectors $X_e \in T_e(G)$ the tangent space to G at the identity element and, in the case of matrix groups, are easily obtainable in terms of the "exponential" of a matrix.

Section 7 is concerned with the set $\mathfrak{X}(M)$ of all C^∞-vector fields on M. It is a vector space over $\boldsymbol{R}$ and a module over the C^∞ functions on M. Moreover it has a naturally defined product, the bracket $[X, Y]$. Using all these ideas one is able to define the Lie derivative $L_X Y$ of the vector field Y in the direction of the field X; this gives a new vector field dependent on X and Y. This derivative results from the group action associated with X, which enables us to compute the change in Y as we move along the orbit.

The final two sections give Frobenius' theorem, a very basic existence theorem in manifold theory, and some applications. These two sections can be omitted on a first reading; they are important but we make relatively few applications of them. For a first reading, in fact, Sections 1-4 are the most crucial.

1 The Tangent Space at a Point of a Manifold

Let M denote a C^∞ manifold of dimension n. We have defined for M the concepts of C^∞ function on an open subset U and of C^∞ mapping to another manifold. This allows us to consider $C^\infty(U)$, the collection of all C^∞ functions on the open subset U (including the special case $U = M$), and to verify—as we did for $U \subset \boldsymbol{R}^n$—that it is a commutative algebra over the real numbers $\boldsymbol{R}$. As before, $\boldsymbol{R}$ may be identified in a natural way with the constant functions and the constant 1 with the unit. Given any point $p \in M$ we may—as for $\boldsymbol{R}^n$—define $C^\infty(p)$ as the algebra of C^∞ functions whose domain of definition includes some open neighborhood of p, with functions identified if they agree on any neighborhood of p. The objects so obtained are called "germs" of C^∞ functions (Exercise 1). Choosing an arbitrary coordinate neighborhood U, φ of p it is easily verified that $\varphi^*: C^\infty(\varphi(p)) \to C^\infty(p)$ given by $\varphi^*(f) = f \circ \varphi$ is an isomorphism of the algebra of "germs" of C^∞ functions at $\varphi(p) \in \boldsymbol{R}^n$ onto the algebra $C^\infty(p)$. This is to be expected since locally M is C^∞-equivalent to $\boldsymbol{R}^n$ by the diffeomorphism φ. Our main purpose is to attach to each $p \in M$ a tangent vector space $T_p(M)$, as was done for $\boldsymbol{R}^n$ and $\boldsymbol{E}^n$. [See Fig. IV.1 for the geometric idea of $T_p(M)$.] Although our first definitions in the latter case giving $T_p(\boldsymbol{R}^n)$ as directed line segments do not generalize, the identification (based on Theorem II.4.1) of $T_p(\boldsymbol{R}^n)$ with directional derivatives does.

(1.1) Definition We define the *tangent space* $T_p(M)$ to M at p to be the set of all mappings $X_p: C^\infty(p) \to \boldsymbol{R}$ satisfying for all $\alpha, \beta \in \boldsymbol{R}$ and $f, g \in C^\infty(p)$ the two conditions

(i) $X_p(\alpha f + \beta g) = \alpha(X_p f) + \beta(X_p g)$ (linearity),

(ii) $X_p(fg) = (X_p f)g(p) + f(p)(X_p g)$ (Leibniz rule),

with the vector space operations in $T_p(M)$ defined by

$$(X_p + Y_p)f = X_p f + Y_p f,$$
$$(\alpha X_p)f = \alpha(X_p f).$$

A *tangent vector to M at p* is any $X_p \in T_p(M)$.

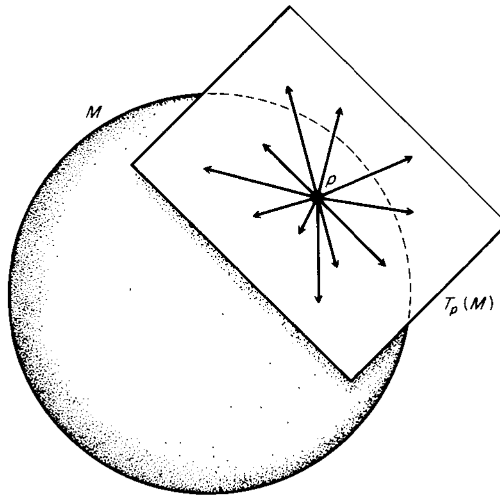


Figure IV.1

One should check that this does in fact define a vector space $T_p(M)$ at each point p of M. Although we are now dealing with C^∞ functions on a manifold, formally the proofs are the same as in Section II.4 where it was established that $\mathscr{D}(a)$, the mappings of $C^\infty(a)$ to $\boldsymbol{R}$ having properties (i) and (ii), was a vector space.

We remark that the definition of $T_p(M)$ uses only $C^\infty(p)$, not all of M; thus if U is any open set of M containing p, then $T_p(U)$ and $T_p(M)$ are *naturally identified*. Of course, our proof that $T_p(M)$ is a vector space includes the earlier case of $\boldsymbol{R}^n$, the difference is that we no longer have the alternative "geometric" way of defining $T_p(M)$ as pairs of points $\overrightarrow{px}$ as we did in $\boldsymbol{R}^n$, because that method used special features of $\boldsymbol{R}^n$, namely the existence of a *natural* one-to-one correspondence with the vector space V^n. For manifolds in general, any such correspondence entails a *choice* of a coordinate neighborhood and depends on the particular neighborhood selected; so it is not natural in the sense we have used the term. However, for each choice of coordinate neighborhood U, φ containing $p \in M$ we obtain an isomorphism to V^n as we shall see. It is by this method that we can establish that $\dim T_p(M) = \dim M$.

(1.2) Theorem *Let $F: M \to N$ be a C^∞ map of manifolds. Then for $p \in M$ the map $F^*: C^\infty(F(p)) \to C^\infty(p)$ defined by $F^*(f) = f \circ F$ is a homomorphism of algebras and induces a dual vector space homomorphism $F_*: T_p(M) \to T_{F(p)}(N)$, defined by $F_*(X_p)f = X_p(F^*f)$, which gives $F_*(X_p)$ as a map of $C^\infty(F(p))$ to $\boldsymbol{R}$. When $F: M \to M$ is the identity, both F^* and F_* are the identity isomorphism. If $H = G \circ F$ is a composition of C^∞ maps, then $H^* = F^* \circ G^*$ and $H_* = G_* \circ F_*$.*

Proof The proof consists of routinely checking the statements against definitions. We omit the verification that F^* is a homomorphism and consider F_* only. Let $X_p \in T_p(M)$ and $f, g \in C^\infty(F(p))$; we must prove that the map $F_*(X_p)$: $C^\infty(F(p)) \to \boldsymbol{R}$ is a vector at $F(p)$, that is, a linear map satisfying the Leibniz rule. We have

$$F_*(X_p)(fg) = X_p F^*(fg) = X_p[(f \circ F)(g \circ F)]$$
$$= X_p(f \circ F)g(F(p)) + f(F(p))X_p(f \circ G),$$

and so we obtain

$$F_*(X_p)(fg) = (F_*(X_p)f)g(F(p)) + f(F(p))F_*(X_p)g$$

(linearity is even simpler). Thus $F_*: T_p(M) \to T_{F(p)}(N)$. Further, F_* is a homomorphism

$$F_*(\alpha X_p + \beta Y_p)f = (\alpha X_p + \beta Y_p)(f \circ F) = \alpha X_p(f \circ F) + \beta Y_p(f \circ F)$$
$$= \alpha F_*(X_p)f + \beta F_*(Y_p)f$$
$$= [\alpha F_*(X_p) + \beta F_*(Y_p)]f. \quad \blacksquare$$

(1.3) Remark The homomorphism $F_*: T_p(M) \to T_{F(p)}(M)$ is often called the *differential* of F. One frequently sees other notations for F_*, for example, dF, DF, F', and so on. The $*$ is a subscript since the mapping is in the same "direction" as F, that is, from M to N, whereas F^*: $C^\infty(F(p)) \to C^\infty(p)$ goes opposite to the direction of F. This notational convention can be quite important and reflects a similar situation in linear algebra related to linear mappings of vector spaces and their duals.

Although, once definitions are correctly made and rather mechanically applied, the statements above have trivial proofs, nonetheless they are most important and useful, even if M and N are Euclidean spaces. We shall consider some of the consequences now.

(1.4) Corollary *If $F: M \to N$ is a diffeomorphism of M onto an open set $U \subset N$ and $p \in M$, then $F_*: T_p(M) \to T_{F(p)}(N)$ is an isomorphism onto.*

This follows at once from the last statement of the theorem and the remark after Definition 1.1 if we suppose G is inverse to F. Then both $G_* \circ F_*: T_p(M) \to T_p(M)$ and $F_* \circ G_*: T_{F(p)}(N) \to T_{F(p)}(N)$ are the identity isomorphism on the corresponding vector space.

Remembering that any open subset of a manifold is a (sub)manifold of the same dimension, we see that if U, φ is a coordinate neighborhood on M, then the coordinate map φ induces an isomorphism $\varphi_*: T_p(M) \to T_{\varphi(p)}(\boldsymbol{R}^n)$ of the tangent space at each point $p \in U$ onto $T_a(\boldsymbol{R}^n)$, $a = \varphi(p)$. The map φ^{-1},

on the other hand, maps $T_a(\boldsymbol{R}^n)$ isomorphically onto $T_p(M)$. The images $E_{ip} = \varphi_*^{-1}(\partial/\partial x^i)$, $i = 1, \ldots, n$, of the natural basis $\partial/\partial x^1, \ldots, \partial/\partial x^n$ at each $a \in \varphi(U) \subset \boldsymbol{R}^n$ determine at $p = \varphi^{-1}(a) \in M$ a basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$; we call these bases the *coordinate frames*.

(1.5) Corollary *To each coordinate neighborhood U on M there corresponds a natural basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$ for every $p \in U$; in particular,* $\dim T_p(M) = \dim M$. *Let f be a C^∞ function defined in a neighborhood of p, and $\hat{f} = f \circ \varphi^{-1}$ its expression in local coordinates relative to U, φ. Then $E_{ip} f = (\partial \hat{f}/\partial x^i)_{\varphi(p)}$. In particular, if $x^i(q)$ is the* ith *coordinate function, $X_p x^i$ is the* ith *component of X_p in this basis, that is, $X_p = \sum_{i=1}^n (X_p x^i) E_{ip}$.*

The last statement of the corollary is a restatement of the definition in Theorem 1.2 for $E_{ip} = \varphi_*^{-1}(\partial/\partial x^i)$, namely,

$$E_{ip} f = \left(\varphi_*^{-1}\left(\frac{\partial}{\partial x^i}\right)\right) f = \frac{\partial}{\partial x^i}(f \circ \varphi^{-1})\bigg|_{x = \varphi(p)}.$$

If we take f to be the ith coordinate function, $f(q) = x^i(q)$ and $X_p = \sum \alpha^j E_{jp}$, then

$$X_p x^i = \sum_j \alpha^j (E_{jp} x^i) = \sum_j \alpha^j \left(\frac{\partial x^i}{\partial x^j}\right)_{\varphi(p)} = \alpha^i.$$

We may use this to derive a standard formula which gives the matrix of the linear map F_* relative to local coordinate systems. Let $F: M \to N$ be a smooth map, and let U, φ and V, ψ be coordinate neighborhoods on M and N with $F(U) \subset V$. Suppose that in these local coordinates F is given by

$$y^i = f^i(x^1, \ldots, x^n), \qquad i = 1, \ldots, m,$$

and that p is a point with coordinates $a = (a^1, \ldots, a^n)$. Then $F(p)$ has y coordinates determined by these functions. Further let $\partial y^j/\partial x^i$ denote $\partial f^j/\partial x^i$.

(1.6) Theorem *Let $E_{ip} = \varphi_*^{-1}(\partial/\partial x^i)$ and $\tilde{E}_{jF(p)} = \psi_*^{-1}(\partial/\partial y^j)$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$, be the basis of $T_p(M)$ and $T_{F(p)}(N)$, respectively, determined by the given coordinates neighborhoods. Then*

$$F_*(E_{ip}) = \sum_{j=1}^m \left(\frac{\partial y^j}{\partial x^i}\right)_a \tilde{E}_{jF(p)}, \qquad i = 1, \ldots, n.$$

In terms of components, if $X_p = \sum \alpha^i E_{ip}$ maps to $F_*(X_p) = \sum \beta^j \tilde{E}_{jF(p)}$, then we have

$$\beta^j = \sum_{i=1}^n \alpha^i \left(\frac{\partial y^j}{\partial x^i}\right)_a, \qquad j = 1, \ldots, m.$$

The partial derivatives in these formulas are evaluated at the coordinates of p: $a = (a^1, \ldots, a^n) = \varphi(p)$.

Proof We have $F_*(E_{ip}) = F_* \circ \varphi_*^{-1}(\partial/\partial x^i)_{\varphi(p)}$ and according to Corollary 1.5, to compute its components relative to $\tilde{E}_{jF(p)}$, we must apply this vector as an operator on $C^\infty(F(p))$ to the coordinate functions y^j

$$F_*(E_{ip})y^j = \left(F_* \circ \varphi_*^{-1}\left(\frac{\partial}{\partial x^i}\right)\right) y^j = \frac{\partial}{\partial x^i} y^j(F \circ \varphi^{-1})(x) = \frac{\partial f^j}{\partial x^i},$$

these derivatives being evaluated at the coordinates of p, that is, at $\varphi(p)$; they could also be written $(\partial y^j/\partial x^i)_{\varphi(p)}$. ∎

We now obtain two corollaries to Theorem 1.6, in the first, F, M, and N are as in Theorem 1.6.

(1.7) Corollary *The rank of F at p is exactly the dimension of the image of $F_*(T_p(M))$. F_* is an isomorphism into if and only if this rank is the dimension of M; it is onto if and only if the rank equals* dim N.

Proof We obtain this immediately from linear algebra since $(\partial y^i/\partial x^j)$ is exactly the Jacobian of $\psi \circ F \circ \varphi^{-1}$, which we used to define the rank in Definition III.4.1, and is also the matrix of the linear transformation $F_*\colon T_p(M) \to T_{F(p)}N$ in the given bases. ∎

This corollary gives a characterization of the rank which is independent of any coordinate systems, a situation toward which we constantly strive in studying properties of various objects on manifolds.

If we apply the theorem to the maps $F = \tilde{\varphi} \circ \varphi^{-1}$ and $F^{-1} = \varphi \circ \tilde{\varphi}^{-1}$ which give the change of coordinates from U, φ to $\tilde{U}, \tilde{\varphi}$ in $U \cap \tilde{U}$ on M, then we obtain formulas for change of basis in $T_p(M)$ and the corresponding change of components relative to these bases.

(1.8) Corollary *Let* $p \in U \cap \tilde{U}$ *and let* $E_{ip} = \varphi_*^{-1}(\partial/\partial x^i)$ *and* $\tilde{E}_{ip} = \tilde{\varphi}_*^{-1}(\partial/\partial \tilde{x}^i)$ *be the bases of $T_p(M)$ corresponding to the two coordinate systems. Then with indices running from* 1 *to n, we have*

$$E_{ip} = \textstyle\sum_k \left(\frac{\partial \tilde{x}^k}{\partial x^i}\right)_{\varphi(p)} \tilde{E}_{kp} \qquad \textit{and} \qquad \tilde{E}_{jp} = \textstyle\sum_l \left(\frac{\partial x^l}{\partial \tilde{x}^j}\right)_{\tilde{\varphi}(p)} E_{lp}\,.$$

If $X_p = \sum \alpha^i E_{ip} = \sum \beta^j \tilde{E}_{jp}$, *then*

$$\alpha^i = \textstyle\sum_j \beta^j \frac{\partial x^i}{\partial \tilde{x}^j} \qquad \textit{and} \qquad \beta^j = \textstyle\sum_i \alpha^i \frac{\partial \tilde{x}^j}{\partial x^i}.$$

The proof is left as an exercise. The second set of formulas is often used to define tangent vector at a point p of a manifold: a tangent vector X_p is an equivalence class of the collection of all n-tuples $\{(\alpha^1, \ldots, \alpha^n)_{(U, \varphi)} | \alpha^i \in \boldsymbol{R}, U, \varphi$ a coordinate neighborhood of $p\}$; two such n-tuples $(\alpha^1, \ldots, \alpha^n)_{U, \varphi}$ and $(\beta^1, \ldots, \beta^n)_{\tilde{U}, \tilde{\varphi}}$ being equivalent if they are related as in last formula of Corollary 1.8 (see Exercise 4).

We may apply Corollary 1.7 to the following situation: M is a *submanifold* of N with $F: M \to N$ the immersion or inclusion map of M into N. In either case, the mapping F from M (with its C^∞ manifold structure) into N (with its C^∞ structure) is a C^∞ mapping, and rank $F = \dim M$. This means that $F_*: T_p(M) \to T_p(N)$ is an injective isomorphism so that $T_p(M)$ can be *identified* with a subspace of $T_p(N)$. This identification being made we can think of $T_p(M)$, the tangent space to M, as a subspace in $T_p(N)$ for each $p \in M$. Applying this principle to our examples of submanifolds of $\boldsymbol{R}^n$, especially when $n = 2$ or 3, will enable us to recapture some of the intuitive meaning of tangent vector which was lost in the transition from Euclidean space to general manifolds. Of course this applies only to those manifolds which can be realized as, that is, are diffeomorphic to, a submanifold of $\boldsymbol{R}^n$.

(1.9) Example Consider the case of a C^∞ curve $F: M \to N$ in a manifold, where $M = (a, b)$ is an open interval of $\boldsymbol{R}$; for the moment we drop the requirement that F is an immersion. Given $t_0 \in M$, $a < t_0 < b$, then d/dt taken at t_0 is a basis for $T_{t_0}(M)$. Suppose $p = F(t_0)$ and $f \in C^\infty(p)$, then $F_*(d/dt)$ is determined by its value on all such f:

$$F_*\left(\frac{d}{dt}\right) f = \left(\frac{d}{dt}(f \circ F)\right)_{t_0}.$$

We shall call this vector the (*tangent*) *velocity vector* to the curve at p. (Fig.IV.2). In this interpretation we use the parameter $t \in \boldsymbol{R}$ as time, and we think of $F(t)$ as a point moving in N.

In particular, if U, φ are coordinates around p, then in the local coordinates F is given by $\hat{F}(t) = \varphi \circ F(t) = (x^1(t), \ldots, x^n(t))$. The ith coordinate x^i is a function on U and using somewhat sloppy notation, we write $x^i(t) = (x^i \circ F)(t)$; thus $F_*(d/dt)x^i = (dx^i/dt)_{t_0}$, which we denote $\dot{x}^i(t_0)$, $i = 1, \ldots, n$. So by Theorem 1.6 (with $E_{1p} = d/dt$ and E's replacing $\tilde{E}$'s),

$$F_*\left(\frac{d}{dt}\right) = \sum_{i=1}^{n} \dot{x}^i(t_0) E_{ip}.$$

Now as a special case let $N = \boldsymbol{R}^n$. With the usual (canonical) coordinates of $\boldsymbol{R}^n$ this formula means that the image of d/dt is just the velocity vector at the point $p = (x^1(t_0), \ldots, x^n(t_0))$ of the curve. Its components relative to the natural basis at the point p are $\dot{x}^1(t_0), \ldots, \dot{x}^n(t_0)$; it is the vector of $T_p(\boldsymbol{R}^n)$

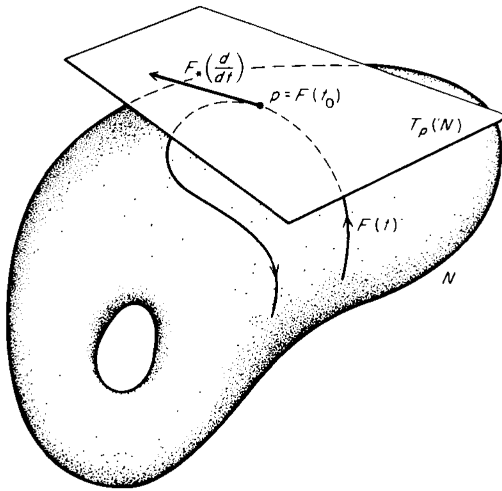


Figure IV.2

whose initial point is $p = x(t_0)$ and whose terminal point is $(x^1(t_0) + \dot{x}^1(t_0), \ldots, x^n(t_0) + \dot{x}^n(t_0))$. If the rank of F at t_0 is 1, then F_* is an isomorphism and we may identify the tangent space to the image curve at p with the subspace of $T_p(\boldsymbol{R}^n)$ spanned by this vector, thus obtaining the usual tangent line at the point p of the curve. If the rank of F at t_0 is 0, then $F_*(d/dt) = 0$.

(1.10) Example We now suppose M to be a two-dimensional submanifold of $\boldsymbol{R}^3$, that is, a surface. Let W be an open subset, say a rectangle in the (u, v)-plane $\boldsymbol{R}^2$ and $\theta\colon W \to \boldsymbol{R}^3$ a parametrization of a portion of M (Fig.IV.3). Namely, suppose θ is an imbedding whose image is an open subset V of M; V, θ^{-1} is a coordinate neighborhood on M. Suppose $\theta(u_0, v_0) = (x_0, y_0, z_0)$, where we now use (x, y, z) as the natural coordinates in $\boldsymbol{R}^3$. We may assume that θ is given by coordinate functions

$$x = f(u, v), \qquad y = g(u, v), \qquad z = h(u, v).$$

Since θ is an imbedding, the Jacobian matrix $\partial(f, g, h)/\partial(u, v)$ has rank 2 at each point of W. We consider the image of the basis vectors $\partial/\partial u$ and $\partial/\partial v$ at (u_0, v_0). We denote these by $(X_u)_0$ and $(X_v)_0$. According to the first formula of Theorem 1.6, they are given by

$$(X_u)_0 = \theta_*\left(\frac{\partial}{\partial u}\right) = \frac{\partial x}{\partial u}\frac{\partial}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial}{\partial y} + \frac{\partial z}{\partial u}\frac{\partial}{\partial z},$$

$$(X_v)_0 = \theta_*\left(\frac{\partial}{\partial v}\right) = \frac{\partial x}{\partial v}\frac{\partial}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial}{\partial z},$$

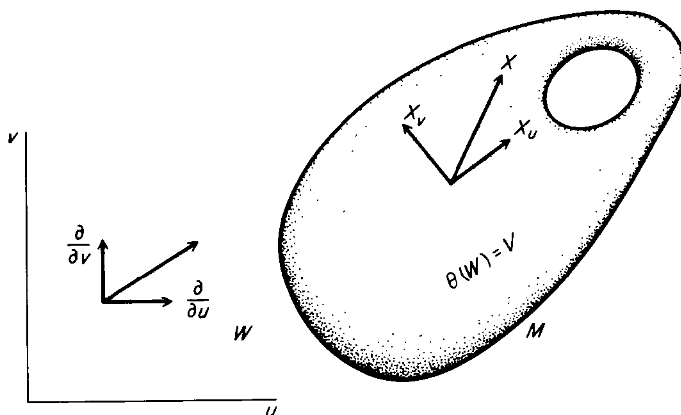


Figure IV.3

where we have written $\partial x/\partial u$, $\partial x/\partial v$ for $\partial f/\partial u$, $\partial f/\partial v$, and so on, these derivatives being evaluated at u_0, v_0. Since θ_* has rank 2, these are linearly independent vectors, and they span a two-dimensional subspace of $T_{(x_0, y_0, z_0)}(\boldsymbol{R}^3)$. This subspace is what we have, by our identification, agreed to call the *tangent space of M at the point* (x_0, y_0, z_0); it consists of all the vectors of the form $\alpha\theta_*(\partial/\partial u) + \beta\theta_*(\partial/\partial v) = \alpha(X_u)_0 + \beta(X_v)_0$, $\alpha, \beta \in \boldsymbol{R}$; their initial point, of course, is always at (x_0, y_0, z_0). It is easily seen that this subspace is the usual tangent plane to a surface, as we would naturally expect it to be. We use one of the standard descriptions of the tangent plane at a point p of a surface M in $\boldsymbol{R}^3$: the collection of all tangent vectors at p to curves through p which lie on M. In fact let I be an open interval about $t = t_0$ and let us consider a curve on N through (x_0, y_0, z_0). It is no loss of generality to suppose the curve given by $F: I \to W$ composed with $\theta: W \to \boldsymbol{R}^3$; thus u, v, are functions of t with $u(t_0) = u_0$ and $v(t_0) = t_0$ and the curve is given by

$$\theta(F(t)) = \big(x(u(t), v(t)), y(u(t), v(t)), z(u(t), v(t))\big).$$

The tangent to the curve at (x_0, y_0, z_0) is given by

$$(\theta \circ F)_*\left(\frac{d}{dt}\right) = \dot{x}(t_0)\frac{\partial}{\partial x} + \dot{y}(t_0)\frac{\partial}{\partial y} + \dot{z}(t_0)\frac{\partial}{\partial z},$$

where

$$\dot{x}(t_0) = \left(\frac{dx}{dt}\right)_{t_0} = \frac{\partial x}{\partial u}\frac{du}{dt} + \frac{\partial x}{\partial v}\frac{dv}{dt}$$

evaluated at (x_0, y_0, z_0) and $t = t_0$. Substituting and collecting terms, we have

$$(\theta \circ F)_*\left(\frac{d}{dt}\right) = \frac{du}{dt}\left(\frac{\partial x}{\partial u}\frac{\partial}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial}{\partial y} + \frac{\partial z}{\partial u}\frac{\partial}{\partial z}\right)$$
$$+ \frac{dv}{dt}\left(\frac{\partial x}{\partial v}\frac{\partial}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial}{\partial z}\right)$$
$$= \frac{du}{dt}\theta_*\left(\frac{\partial}{\partial u}\right) + \frac{du}{dt}\theta_*\left(\frac{\partial}{\partial v}\right)$$
$$= \dot{u}(t_0)(X_u)_0 + \dot{v}(t_0)(X_v)_0.$$

If we let $u = t$, $v = v_0$, we obtain just $(X_u)_0 = \theta_*(\partial/\partial u)$ and analogously $(X_v)_0$ is tangent to the parameter curve $u = u_0$, $v = t$. *The coordinate frame vectors are tangent to the coordinate curves.*

This could also be derived directly from the relation $(\theta \circ F)_* = \theta_* \circ F_*$ of Theorem 1.2. This means that the (tangent) velocity to every curve in M through $p = (x_0, y_0, z_0)$ lies in the subspace $T_p(M) \subset T_p(\boldsymbol{R}^3)$ spanned by $(X_u)_0$ and $(X_v)_0$. Conversely by suitable choice of the curve every vector of $T_p(M)$ may be so represented.

Exercises

1. Let $\mathscr{F}_p$ be the family of C^∞ functions f on open sets W_f of M which contain the point p. Define a relation $\sim$ on $\mathscr{F}_p$ by $f \sim g$ if $f \equiv g$ on a neighborhood of p. Show that $\sim$ is an equivalence relation and that the equivalence classes, called *germs* of C^∞ functions at p, form an algebra $C(p)$ with unit over $\boldsymbol{R}$.
2. Let $F: M \to N$ be a C^∞ mapping of manifolds. Show that $f \to F^*(f) = F \circ f$ defines a homomorphism $F^*: C^\infty(F(p)) \to C^\infty(p)$ and prove the statements of Theorem 1.2 about F^*.
3. Prove Corollary 1.8.
4. For $p \in M$ let $\mathscr{C}$ be the collection of *all* coordinate neighborhoods containing p. Let $(\alpha_1, \ldots, \alpha_n)_{U, \varphi}$ and $(\beta_1, \ldots, \beta_n)_{\tilde{U}, \tilde{\varphi}}$ be objects consisting of an element of $\boldsymbol{R}^n$ together with—or labeled by—a coordinate neighborhood of $\mathscr{C}$. They will be called *equivalent* if they correspond by the formulas of Corollary 1.8. Show that this is an equivalence relation and that the classes form a vector space naturally isomorphic to $T_p(M)$.
5. Using the notation of Example 1.10, show that for any $\alpha, \beta \in \boldsymbol{R}$ there is a parametrized curve on M through p whose velocity vector is exactly $\alpha(X_u)_0 + \beta(X_v)_0$.
6. If the surface of Example 1.10 is given in the form $z = h(x, y)$ with $z_0 = h(x_0, y_0)$, then show as a special case of our discussion that with

suitable parametrization the tangent plane $T_{(x_0, y_0, z_0)}(M)$ consists of all vectors from (x_0, y_0, z_0) to points (x, y, z) satisfying

$$\left(\frac{\partial h}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial h}{\partial y}\right)_0 (y - y_0) - (z - z_0) = 0.$$

7. Let $N \subset M$ be a regular submanifold and U, φ be a preferred coordinate neighborhood relative to N with local coordinates $(x^1, \dots, x^m)$ and frames $(E_1, \dots, E_m)$. If $N \cap U$ is given by $x^{n+1} = \cdots = x^m = 0$, show that $E_{1p}, \dots, E_{np}$ is a basis of $T_p(N)$ for every $p \in N \cap U$. Modify this statement so as to include immersed submanifolds.

2 Vector Fields

In a previous paragraph (Definition 1.1), we defined the notion of a tangent vector to a manifold at a point $p \in M$, that is, an element X_p of $T_p(M)$. In this section we will define and give examples of a C^r-vector field on M, $r \geq 0$. A vector field X on M is, first of all, a "function" assigning to each point p of M an element X_p of $T_p(M)$ (see Fig. IV.4). We place the word "function" in quotation marks since we have not really defined its range, only its domain M. The range is, in fact, the set $T(M)$ consisting of all tangent vectors at all points of M, $T(M) = \bigcup_{p \in M} T_p(M)$.

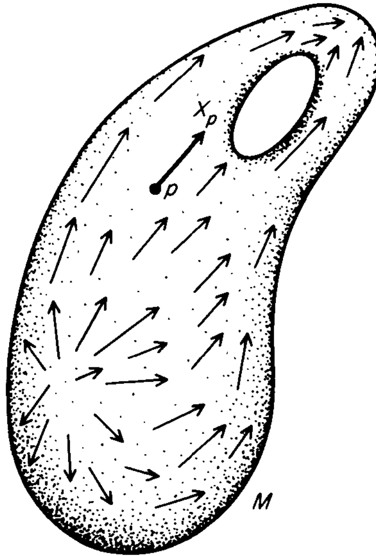


Figure IV.4
Vector field X on M.

For future reference we note some properties of $T(M)$. It is a set that is partitioned into disjoint subsets $\{T_p(M)\}$ which are indexed by the points of M, that is, to $p \in M$ corresponds its tangent space $T_p(M)$. It follows that there is a natural projection $\pi: T(M) \to M$ taking the vector $X_p \in T(M)$ to p. The vector field X as a function $X: M \to T(M)$, must satisfy the condition $\pi \circ X = i_M$, the identity on M. Further details are given in Exercises 5–7.

Second, a vector field X is required to satisfy some condition of regularity, that is, of continuity or differentiability. We impose this as follows: For $p \in M$ let U, φ be any coordinate neighborhood of p, and let $E_{1p}, \ldots, E_{np}$ be the corresponding basis (coordinate frames) of $T_p(M)$. Then X_p, the value of X at p, may be written uniquely as $X_p = \sum_{i=1}^n \alpha^i E_{ip}$. If p is varied in U, the components $\alpha^1, \ldots, \alpha^n$ are well-defined functions of p which must, then, be given by functions of the local coordinates (denoted by the same letters)

$$\alpha^i = \alpha^i(x^1, \ldots, x^n), \qquad i = 1, \ldots, n, \quad \text{on} \quad \varphi(U) \subset \boldsymbol{R}^n.$$

We say that X is of *class* C^r, $r \geq 0$, if these functions are of class C^r on U for every local coordinate system U, φ. Since the expressions given in Corollary 1.8 (see also Exercise 4) are linear with C^∞ coefficients, we see that this definition is independent of the coordinates used. (Note that we include the case $r = 0$ of continuous components.) Collecting these requirements leads to the precise definition:

(2.1) Definition *A vector field X of class C^r on M* is a function assigning to each point p of M a vector $X_p \in T_p(M)$ whose components in the frames of any local coordinates U, φ are functions of class C^r on the domain U of the coordinates. Unless otherwise noted we will use *vector field* to mean C^∞-vector field hereafter.

We remark that this definition is somewhat awkward, especially as regards the regularity condition; our treatment places reliance on local coordinates. One way to avoid this is to define X to be C^r if for every C^∞ function f whose domain W_f is an open subset of U, the function Xf, defined by $(Xf)(p) = X_p f$, is of class C^r. Another very elegant approach is to give $T(M)$ the structure of a C^∞ manifold and then X becomes a mapping, $X: M \to T(M)$, of one C^∞ manifold to another. In this case we have already defined the meaning of C^r in Definition III.3.3. We shall develop these important ideas in the exercises.

(2.2) Example If we consider $M = \boldsymbol{R}^3 - \{0\}$, then the gravitational field of an object of unit mass at 0 is a C^∞-vector field whose components $\alpha^1, \alpha^2, \alpha^3$ relative to the basis $\partial/\partial x^i = E_1$, $\partial/\partial x^2 = E_2$, and $\partial/\partial x^3 = E_3$ are

$$\alpha^i = \frac{x^i}{r^3}, \qquad i = 1, 2, 3 \quad \text{with} \quad r = ((x^1)^2 + (x^2)^2 + (x^3)^2)^{1/2}.$$

(2.3) Example Given any coordinate neighborhood U, φ on a manifold M, then U, being an open set of a manifold, is itself a manifold of the same dimension, say n. The vector fields $E_i = \varphi_*^{-1}(\partial/\partial x^i)$, $i = 1, \ldots, n$, have components $\alpha^j = \delta_i^j$. These are constants and hence C^∞ functions on U, so that each E_i is a C^∞-vector field on U. The set $E_1, \ldots, E_n$ is a basis of $T_p(M)$ at each $p \in U$, the *coordinate frames* (Fig. IV.5).

More generally, a set of k vector fields on a manifold M, $\dim M = n$, which is linearly independent at each point is called a *field of k-frames on M*. If $k = n$, then the frames form a basis at each point. Of course, it would be convenient if on a manifold one could always find such a field of n-frames,

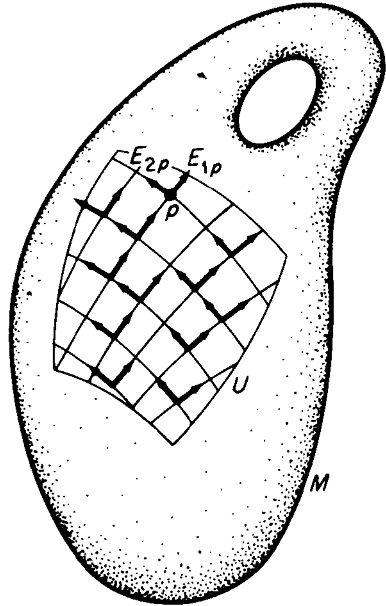


Figure IV.5
Coordinate frames on $U \subset M$.

for then the components of any vector field would be globally defined, that is, functions whose domain is all of M. This would relieve us of the necessity of using local coordinate neighborhoods and the associated frames $E_1, \ldots, E_n$. However, it is known that this is not possible in general, for example, on the sphere S^2 it is not possible to define even one continuous vector field X which is linearly independent (nonzero) at each point of S^2. This a classical theorem of algebraic topology discovered by Brouwer; it will be proved in Section VI.8. We shall give some further related examples for which we need the following lemma:

(2.4) Lemma *Let N be a submanifold of M and let X be a C^∞-vector field on M such that for each $p \in N$, $X_p \in T_p(N)$. Then X restricted to N is a C^∞-vector field on N (called the restriction of X to N).*

Proof By hypothesis X assigns to each $p \in N$ the tangent vector X_p in the subspace $T_p(N)$ of $T_p(M)$. We must prove that X restricted to N is of class C^∞. By choosing suitable cubical coordinate neighborhoods V, ψ and U, φ of $p \in N$ in N and in M, respectively (cf. III.4.13 and III.5.6), we may assume that the immersion (inclusion) is given locally by $i:(y^1, \ldots, y^n) \to (y^1, \ldots, y^n, 0, \ldots, 0)$. Thus V is a slice through p of U and $\psi = \pi \circ \varphi | V$, π denoting projection to the first n coordinates. We have $\psi^{-1} = \varphi^{-1} \circ i$ and hence for $q \in V$, $\psi_*^{-1}(\partial/\partial y_i) = \varphi_*^{-1}(\partial/\partial x_i)$, so the coordinate frames at $q \in V$ are naturally identified with first n vectors $E_{1q}, \ldots, E_{nq}$ of the coordinate frames on U, $E_{iq} = \varphi_*^{-1}(\partial/\partial x_i)$. Now if on U, the C^∞ vector field $X_q = \sum_{i=1} \alpha_i(q)E_{iq}$, then for $q \in V$, the last $m - n$ components vanish, since for $q \in V$, $X_q \in T_q(N)$ by hypothesis. Thus, restricted to V, $X_q = \sum_{i=1}^n \alpha_i(q)E_{iq}$, where $\alpha_1(q), \ldots, \alpha_n(q)$ are the restrictions to the submanifold V of U and therefore C^∞ on V. This shows that the restriction to N of X is C^∞ in a neighborhood V of any $p \in N$ and is thus C^∞ on N. ∎

(2.5) Example Although no nonvanishing continuous vector field exists on the 2-sphere S^2, there are *three* mutually perpendicular unit vector fields on $S^3 \subset \mathbf{R}^4$, that is, a frame field. Let $S^3 = \{(x^1, x^2, x^3, x^4) \mid \sum_{i=1}^4 (x^i)^2 = 1\}$ and let the vector fields be given by

$$X = -x^2 \frac{\partial}{\partial x^1} + x^1 \frac{\partial}{\partial x^2} + x^4 \frac{\partial}{\partial x^3} - x^3 \frac{\partial}{\partial x^4},$$

$$Y = -x^3 \frac{\partial}{\partial x^1} - x^4 \frac{\partial}{\partial x^2} + x^1 \frac{\partial}{\partial x^3} + x^2 \frac{\partial}{\partial x^4},$$

$$Z = -x^4 \frac{\partial}{\partial x^1} + x^3 \frac{\partial}{\partial x^2} - x^2 \frac{\partial}{\partial x^3} + x^1 \frac{\partial}{\partial x^4},$$

at the point $x = (x^1, x^2, x^3, x^4)$ of S^3. Since at each point these are mutually orthogonal unit vectors in $\mathbf{R}^4$, they are independent. To see that they are tangent to S^3 it is enough to see that they are orthogonal to the radius vector from the origin 0 to the point x of S^3; this is easy to check. There remains only the question of whether they are C^∞-vector fields. However, this is an immediate consequence of the preceding lemma with $N = S^3$ and $M = \mathbf{R}^4$.

It is possible to show that all odd-dimensional spheres have at least one nonvanishing C^∞-vector field and that—like S^2—no even-dimensional sphere has any continuous nonvanishing field of tangent vectors. It has recently been proved that only the spheres S^1, S^3, S^7 have a C^∞ field of bases as we have just seen to be the case for S^3. Manifolds with this very special property are called *parallelizable*. As already mentioned, coordinate neighborhoods are parallelizable.

Having established the concept of vector field on a manifold, we must now consider what happens when we map a manifold N on which a vector

field is defined into another manifold M. In Section 1 we saw that if $F: N \to M$ is a C^∞ map, then to each point $p \in M$ there is associated a homomorphism $F_*: T_p(N) \to T_{F(p)}(M)$. If X is a vector field on N, then $F_*(X_p)$ is a vector at $F(p)$. However, this process does not in general induce a vector field on M for various reasons: first, $F(N)$ may not be all of M, that is, given $q \in M$ it may well happen that for no $p \in N$ is $F(p) = q$. Second, even if $F^{-1}(q)$ is not empty, it may contain more than one element, say p_1, p_2 with $p_1 \neq p_2$, and then it may happen that $F_*(X_{p_1}) \neq F_*(X_{p_2})$ so that there is no uniquely determined vector Y_q at q which is the image of vectors of the field X on N. It is easy to construct examples of these mishaps, for instance, let N be the half-space $x^1 > 0$ in $\boldsymbol{R}^3$ and $F: N \to M$ be projection to the coordinate plane $x^3 = 0$. If X is the gravitational field of Example 2.2 restricted to N, we see that the image vectors do not determine a vector field on M.

(2.6) Definition If, using the notation above, we have a vector field Y on M such that for each $q \in M$ and $p \in F^{-1}(q) \subset N$ we have $F_*(X_p) = Y_q$, then we say that the vector fields X and Y are *F-related* and we write, briefly, $Y = F_*(X)$. [We do not require F to be onto: If $F^{-1}(q)$ is empty, then the condition is vacuously satisfied.]

(2.7) Theorem *If $F: N \to M$ is a diffeomorphism, then each vector field X on N is F-related to a uniquely determined vector field Y on M.*

Proof Since F is a diffeomorphism, it has an inverse $G: M \to N$, and at each point p we have $F_*: T_p(N) \to T_{F(p)}(M)$ is an isomorphism onto with G_* as inverse. Thus given a C^∞-vector field X on N, then at each point q of M, the vector $Y_q = F_*(X_{G(q)})$ is uniquely determined. It then remains to check that Y is a C^∞-vector field. This is immediate if we introduce local coordinates and apply Theorem 1.6 to the component functions. ∎

We remark that under the hypotheses of Lemma 2.4 we have a second example of F-related vector fields: Let $F: N \to M$ be the inclusion map and let X' be X restricted to N. Then X' and X are F-related by the lemma. Further examples of F-related vector fields arise from the study of Lie groups.

(2.8) Definition If $F: M \to M$ is a diffeomorphism and X is a C^∞ vector field on M such that $F_*(X) = X$, that is, X is F-related to itself, then X is said to be *invariant with respect to F*, or F-invariant.

(2.9) Theorem *Let G be a Lie group and $T_e(G)$ the tangent space at the identity. Then each $X_e \in T_e(G)$ determines uniquely a C^∞-vector field X on G which is invariant under left translations. In particular, G is parallelizable.*

Proof To each $g \in G$ there corresponds exactly one left translation L_g taking e to g. Therefore if it exists, X is uniquely determined by the formula: $X_g = L_{g*}(X_e)$. Except for differentiability, this formula does define a left invariant vector field since for $a \in G$, we have $L_{a*}(X_g) = L_{a*} \circ L_{g*}(X_e) = L_{ag*}(X_e) = X_{ag}$. We must show that X, so determined, is C^∞. Let U, φ be a coordinate neighborhood of e such that $\varphi(e) = (0, \ldots, 0)$ and let V be a neighborhood of e satisfying $VV \subset U$. Let $g, h \in V$ with coordinates $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$, respectively, and let $z = (z^1, \ldots, z^n)$ be the coordinates of the product gh. Then $z^i = f^i(x, y)$, $i = 1, \ldots, n$, are C^∞ functions on $\varphi(V) \times \varphi(V)$. If we write $X_e = \sum_{i=1}^n \gamma^i E_{ie}$, $\gamma^1, \ldots, \gamma^n$ real numbers, then according to Theorem 1.6 the formula above for X_g becomes

$$X_g = L_{g*}(X_e) = \sum \gamma^j \left(\frac{\partial f^i}{\partial y^j}\right)_{(x, 0)} E_{ig}$$

since in local coordinates L_g is given by $z^i = f^i(x, y)$, $i = 1, \ldots, n$, with the coordinates x of g fixed. It follows that on V the components of X_g in the coordinate frames are C^∞ functions of the local coordinates. However, for any $a \in G$ the open set aV is the diffeomorphic image by L_a of V. Moreover X, as noted above, is L_a-invariant so that for every $g = ah \in aV$ we have $X_g = L_{a*}(X_h)$. It follows that X on aV is L_a-related to X on V and therefore X is C^∞ on aV by Theorem 2.7. Since X is C^∞ in a neighborhood of each element of G, it is C^∞ on G. ∎

(2.10) Corollary *Let G_1 and G_2 be Lie groups and $F: G_1 \to G_2$ a homomorphism. Then to each left-invariant vector field X on G_1 there is a uniquely determined left-invariant vector field Y on G_2 which is F-related to X.*

Proof By Theorem 2.9, X is determined by X_{e_1}, its value at the identity e_1 of G_1. Let $e_2 = F(e_1)$ be the identity of G_2 and let Y be the uniquely determined left-invariant vector field on G_2 such that $Y_{e_2} = F_*(X_{e_1})$. That Y should have this value at e_2 is surely a necessary condition for Y to be F-related to X; and it remains only to see whether this vector field Y satisfies $F_*(X_g) = Y_{F(g)}$ for every $g \in G_1$. If so, Y is indeed F-related (and uniquely determined). We write the mapping F as a composition $F = L_{F(g)} \circ F \circ L_{g^{-1}}$, using $F(x) = F(g)F(g^{-1}x)$, and note that since both X and Y are left-invariant by assumption, this gives

$$F_*(X_g) = L_{F(g)*} \circ F_* \circ L_{g^{-1}*}(X_g),$$

$$F_*(X_g) = L_{F(g)*} \circ F_*(X_{e_1}) = L_{F(g)*} Y_{e_1},$$

$$F_*(X_g) = Y_{F(g)}.$$

Therefore Y meets all conditions and the corollary is true. ∎

Exercises

1. Show that a function X assigning to each $p \in M$ an element of $T_p(M)$—as in Definition 2.1—is C^∞ if and only if whenever f is a C^∞ function on an open set W_f of M, then Xf, defined by $(Xf)(p) = X_p f$, is C^∞ on W_f.
2. Show that a C^∞-vector field X on M defines a derivation on $C^\infty(M)$ by $f \to Xf$ as defined in Exercise 1.
3. Show that the derivations of $C^\infty(M)$, M a C^∞ manifold, are in a natural one-to-one correspondence with $\mathfrak{X}(M)$ the collection of all C^∞-vector fields on M.
4. Show that the collection $\mathfrak{X}(M)$ of all C^∞-vector fields on M, is closed under addition and multiplication by C^∞ functions [both defined pointwise: $(X + Y)_p = X_p + Y_p$ and $(fX)_p = f(p)X_p$].
5. Define a C^∞ structure of a manifold on $T(M)$ in such a manner that for each coordinate system U, φ on M, with local coordinates $(x^1, \ldots, x^n)$ and frames $E_1, \ldots, E_n$, the set $\tilde{U} = \pi^{-1}(U)$ with mapping $\tilde{\varphi}: \tilde{U} \to \boldsymbol{R}^{2n} = \boldsymbol{R}^n \times \boldsymbol{R}^n$ defined as follows is a coordinate neighborhood: For $p \in U$, $X_p \in \tilde{U}$, we suppose $X_p = \sum_{i=1}^n \alpha^i E_{ip}$ and define $\tilde{\varphi}(X_p) = (x^1(p), \ldots, x^n(p); \alpha^1, \ldots, \alpha^n) = (\varphi(p); \alpha^1, \ldots, \alpha^n)$.
6. Using Exercise 5, show that $\pi: T(M) \to M$ is C^∞, a submersion, and that $T_p(M) = \pi^{-1}(p)$ is a submanifold of $T(M)$.
7. Using Exercise 5, show that the C^∞-vector fields on M correspond precisely to the C^∞ mappings $X: M \to T(M)$ satisfying $\pi \circ X = i_M$, the identity map on M [$\pi(X_p) = p$)].
8. Suppose $F: N \to M$ is C^∞. Then, in the notation of Exercise 7, show that $F_*: T(N) \to T(M)$ is C^∞ and $\pi \circ F_* = F \circ \pi$.
9. Show that if $F: N \to M$ is C^∞ and X is a C^∞-vector field on N, then an F-related vector field Y on M, if it exists, is *uniquely* determined if and only if $F(N)$ is dense in M. Note that this applies in particular to vector fields on submanifolds, i.e., for $F: N \to M$ a one-to-one immersion.
10. Show that the restriction of
$$Y = x^2 \frac{\partial}{\partial x^1} - x^1 \frac{\partial}{\partial x^2} + \cdots + x^{2n} \frac{\partial}{\partial x^{2n-1}} - x^{2n-1} \frac{\partial}{\partial x^{2n}}$$
on $\boldsymbol{R}^{2n}$ to S^{2n-1} defines a nonvanishing C^∞-vector field on S^{2n-1}.
11. Let $F: \tilde{M} \to M$ be a C^∞ covering and Y any C^∞-vector field on M. Show that there is a unique C^∞-vector field X on $\tilde{M}$ such that X and Y are F-related.
12. Show that any C^∞-vector field Y on $S^{n-1} \subset \boldsymbol{R}^n$ is the restriction of a C^∞-vector field X on $\boldsymbol{R}^n$.

13. Given a C^∞ mapping $F: N \to M$ and C^∞-vector fields X on N and Y on M, show that Y is F-related to X if and only if for any C^∞ function g on M we have $(Yg) \circ F = X(g \circ F)$ on the inverse image $F^{-1}(W_g)$ of the domain W_g of g.

3 One-Parameter and Local One-Parameter Groups Acting on a Manifold

We shall now subject the case of a connected Lie group of dimension 1 acting on a manifold M to the same scrutiny as we did the case of a Lie group of dimension 0 in Section III.8, but with very different emphasis. At that time we were interested in the space of orbits; in the present instance we are mainly concerned with the relation to vector fields on M. For this reason we shall limit ourselves to the action of R, by which we denote the *additive (Lie) group* of real numbers $\boldsymbol{R}$, acting on M since this will illustrate all the relevant facts—it can be shown that R and S^1 are the only connected Lie groups of dimension 1. These two cases, discrete Lie groups and the one-dimensional Lie group R acting on M, will give some idea of the depth and diversity of the whole subject of group action on manifolds. Later we shall have something to say about another special case: transitive action of a Lie group G on a manifold.

Thus, in the present section, we consider Definition III.7.1 specialized to an action θ of R on M. Let $\theta: R \times M \to M$ be a C^∞ mapping which satisfies the two conditions:

(i) $\theta_0(p) = p$ for all $p \in M$,
(ii) $\theta_t \circ \theta_s(p) = \theta_{t+s}(p) = \theta_s \circ \theta_t(p)$ for all $p \in M$ and $s, t \in R$.

[We will often write $\theta(t, p)$ as $\theta_t(p)$ or $\theta_p(t)$, depending on which variable is to be emphasized.]

(3.1) Example Suppose that $M = \boldsymbol{R}^3$ and $a = (a^1, a^2, a^3)$ is fixed and assumed different from 0. Then $\theta_t(x) = (x^1 + a^1 t,\ x^2 + a^2 t,\ x^3 + a^3 t)$ defines a C^∞ action of R on M. To each $t \in R$ we have thus assigned the translation $\theta_t: \boldsymbol{R}^3 \to \boldsymbol{R}^3$, taking the point x to the point $x + ta$. This is a free action and the orbits consist of straight lines parallel to the vector a. A particularly simple special case is given by $a = (1, 0, 0)$ so that $\theta_t(x) = (x^1 + t, x^2, x^3)$.

Suppose that $\theta: R \times M \to M$ is any such C^∞ action. Then it defines on M a C^∞-vector field X, which we shall call the *infinitesimal generator* of θ, according to the following prescription: For each $p \in M$ we define $X_p: C^\infty(p) \to \boldsymbol{R}$ by

(3.2) $$X_p f = \lim_{\Delta t \to 0} \frac{1}{\Delta t}[f(\theta_{\Delta t}(p)) - f(p)].$$

We may check directly from (3.2) that X_p is a vector at p in the sense of Definition 1.1, and then verify that $p \to X_p$ defines a vector field, or we may proceed as follows. Let U, φ be a coordinate neighborhood of $p \in M$ and let $I_\delta \times V$ be an open subset of $(0, p)$ in $R \times M$, where $I_\delta = \{t \in R \mid -\delta < t < \delta\}$, and V and $\delta > 0$ are so chosen that $\theta(I_\delta \times V) \subset U$. In particular, $V = \theta_0(V)$ is contained in U and contains p. Restricted to the open set $I_\delta \times V$, we may write θ in local coordinates

$$\begin{aligned} y^1 &= h^1(t, x^1, \ldots, x^n), \\ &\vdots \\ y^n &= h^n(t, x^1, \ldots, x^n), \end{aligned}$$

or $y = h(t, x)$, where $x = (x^1, \ldots, x^n)$ are the coordinates of $q \in V$ and $y = (y^1, \ldots, y^n)$ of $\theta_t(q)$, its image. The h^i are defined and C^∞ on $I_\delta \times \varphi(V)$ and the range of $h(t, x)$ is in $\varphi(U)$. The fact that θ_0 is the identity and $\theta_{t_1 + t_2} = \theta_{t_1} \circ \theta_{t_2}$ is reflected in the conditions:

$$h^i(0, x) = x^i \qquad \text{and} \qquad h^i(t_1 + t_2, x) = h^i(t_1, h(t_2, x))$$

for $i = 1, \ldots, n$. Now if $\hat{f}(x^1, \ldots, x^n)$ is the local expression for $f \in C^\infty(p)$, then

$$\frac{1}{\Delta t}[f(\theta_{\Delta t}(p)) - f(p)] = \frac{1}{\Delta t}[\hat{f}(h(\Delta t, x)) - \hat{f}(x)]$$

and

$$X_p f = \lim_{\Delta t \to 0} \frac{1}{\Delta t}[\hat{f}(h(\Delta t, x)) - \hat{f}(x)] = \sum_{i=1}^{n} \dot{h}^i(0, x)\left(\frac{\partial \hat{f}}{\partial x^i}\right)_{\varphi(p)},$$

where we have used a dot to indicate differentiation with respect to t. This formula is valid for every $p \in V$ and implies that on V, $X_p = \sum \dot{h}^i(0, x) E_{ip}$ with $E_i = \varphi_*^{-1}(\partial/\partial x^i)$ and $x = \varphi(p)$, which shows that X is a C^∞-vector field over V. Since every point of M lies in such a neighborhood, X is C^∞ on M. Note that definition of X at $p \in M$ involves only the values of θ on $I_\delta \times V$, that is, like derivatives in general, it is defined locally and involves only values of t near $t = 0$.

(3.3) Definition If $\theta\colon G \times M \to M$ is the action of a group G on a manifold M, then a vector field X on M is said to be *invariant under the action of* G or *G-invariant* if X is invariant under each of the diffeomorphisms θ_g of M to itself, in brief if $\theta_{g*}(X) = X$ (as in Definition 2.8).

(3.4) Theorem *If $\theta\colon R \times M \to M$ is a C^∞ action of R on M, then the infinitesimal generator X is invariant under this action, that is, $\theta_{t*}(X_p) = X_{\theta_t(p)}$ for all $t \in R$.*

Proof Let $f \in C^\infty(\theta_t(p))$ for some $(t, p) \in R \times M$ and compute $\theta_{t*}(X_p)f$:

$$\theta_{t*}(X_p)f = X_p(f \circ \theta_t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [f \circ \theta_t(\theta_{\Delta t}(p)) - f \circ \theta_t(p)].$$

However, R is Abelian and we have $\theta_t \circ \theta_{\Delta t} = \theta_{t+\Delta t} = \theta_{\Delta t} \circ \theta_t$ so

$$\theta_{t*}(X_p)f = \lim_{\Delta t \to 0} \frac{1}{\Delta t}[(f \circ \theta_{\Delta t})(\theta_t(p)) - f(\theta_t(p))] = X_{\theta_t(p)} f.$$

Since this holds for all f, the result follows. ∎

(3.5) Corollary *If $X_p = 0$, then for each q in the orbit of p we have $X_q = 0$, that is, at the points of an orbit the associated vector field vanishes identically or is never zero.*

Proof The orbit of p consists of all q such that $q = \theta_t(p)$ for some $t \in R$; thus by the theorem $X_q = \theta_{t*} X_p$. Since θ_t is a diffeomorphism, we know that θ_{t*} is an isomorphism of $T_p(M)$ onto $T_q(M)$ so that $X_q = 0$ if and only if $X_p = 0$. ∎

(3.6) Theorem *The orbit of p is either a single point or an immersion of R in M by the map $t \to \theta_t(p)$, depending on whether or not $X_p = 0$.*

Proof The orbit of p is the image of R under the C^∞ map $t \to \theta_t(p)$ into M. Denote this map by F so that $F(t) = \theta_t(p)$. Let $t_0 \in R$ and d/dt denote the standard basis of $T_{t_0}(R)$; F is an immersion if and only if $F_*(d/dt) \neq 0$ for every $t_0 \in R$. Let $f \in C^\infty(F(t_0)) = C^\infty(\theta_{t_0}(p))$ and observe that

$$\begin{aligned} F_*\left(\frac{d}{dt}\right) f = \frac{d}{dt}(f \circ F)_{t_0} &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}[f \circ F(t_0 + \Delta t) - f \circ F(t_0)] \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}[f(\theta_{t_0+\Delta t}(p)) - f(\theta_{t_0}(p))] \\ &= X_{\theta_{t_0}(p)} f \end{aligned}$$

by precisely the same arguments as we used to prove Theorem 3.4. This formula and Corollary 3.5 show that either $X_p \neq 0$ and F is an immersion or else $X_{F(t)} = F_*(d/dt) \equiv 0$, in which case F is a constant map with $F(R) = p$. For the proof of this last statement see Exercise 1. ∎

We remark that the formula just obtained, namely,

$$F_*\left(\frac{d}{dt}\right) = X_{\theta_{t_0}(p)} = X_{F(t_0)},$$

shows that at each point $p \in M$ the vector X_p is tangent to its orbit and in fact is the (tangent) velocity vector of the curve $t \to F(t)$ in M in the sense in

which we have previously (Example 1.9) defined the velocity vector to a parametrized curve, that is, to a differentiable map of an open interval J of R into M, namely $F_*(d/dt)$. This latter notation is not too precise since it does not indicate that $d/dt \in T_{t_0}(R)$ and that F_* is a homomorphism of this one-dimensional vector space into $T_{F(t_0)}(M)$. For this reason we often will write either $\dot{F}(t_0)$ or $(dF/dt)_{t_0}$ to denote the velocity vector. Sometimes it is convenient to let $t \to p(t)$ denote the mapping rather than F. Then its velocity vector is written dp/dt or $\dot{p}(t)$. For example, in the notation of Theorem 3.6, the formula above can be written $\dot{\theta}(t, p) = X_{\theta(t, p)}$.

If we change parameter by a function $t = f(s)$ so that $s \to G(s) = F(f(s))$ gives the curve, then for $t_0 = f(s_0)$,

$$\left(\frac{dG}{ds}\right)_{s_0} = G_*\left(\frac{d}{ds}\right) = F_* \circ f_*\left(\frac{d}{ds}\right) = F_*\left(\frac{dt}{ds}\frac{d}{dt}\right),$$

which give the formula

$$\left(\frac{dG}{ds}\right)_{s_0} = \left(\frac{dt}{ds}\right)_{s_0} F_*\left(\frac{d}{dt}\right)_{t_0}.$$

Thus the velocity vector with respect to s is a scalar multiple by $(dt/ds)_{s_0}$ of the velocity vector with respect to t. This may be conveniently written

$$\text{(3.7)} \qquad \dot{G} = \left(\frac{dt}{ds}\right)\dot{F}(f(s)) \qquad \text{or} \qquad \frac{dp}{ds} = \frac{dp}{dt}\frac{dt}{ds}.$$

This vector equation is, of course, just a special case of the chain rule.

(3.8) Definition Given a vector field X on a manifold M, we shall say that a curve $t \to F(t)$ defined on an open interval J of $\boldsymbol{R}$ is an *integral curve* of X if $dF/dt = X_{F(t)}$ on J. By definition, an integral curve is connected.

We have just shown that each orbit of the action θ is an integral curve of X, the infinitesimal generator of θ, that is, for each fixed $p \in M$, $\dot{\theta}(t, p) = X_{\theta(t, p)}$.

At this point some natural questions arise concerning vector fields and one-parameter group actions: Is every C^∞-vector field the infinitesimal generator of some group action? Can two different actions of R on M give rise to the same vector field X as infinitesimal generator? These questions will be answered in this and the next section. However, a simple but instructive example will illustrate the difficulties we face and show the necessity for a less restrictive concept of one-parameter group action.

(3.9) Example Let $M = \boldsymbol{R}^2$ and let $\theta: R \times M \to M$ be defined by $\theta(t, (x, y)) = (x + t, y)$. Then the infinitesimal generator is $X = \partial/\partial x$. This

action is given by translation of each point (x, y) to a point t units to the right. Suppose now that we remove the origin $(0, 0)$ from $\boldsymbol{R}^2$; let $M_0 = \boldsymbol{R}^2 - \{(0, 0)\}$. For most points θ_t is defined as before; however, we cannot obtain an action of R on M_0 by restriction of θ to $R \times M_0$ since points of the closed set $F = \{(t, (x, 0)) \mid t + x = 0\} = \theta^{-1}(0, 0)$ of $R \times M$ are mapped by θ to the origin. Let $W \subset R \times M_0$ be the open set

$$W = \left(\bigcup_{y \neq 0} R \times \{(x, y)\}\right) \cup \left(\bigcup_{y \neq 0} \{t \in R \mid x(x + t) > 0\} \times \{(x, 0)\}\right).$$

Then $\theta = \theta \mid W$ maps W onto M_0 and preserves many of the features of θ which we have used. For example, let $p = (x, y) \in M_0$, then

(i) $(0, p) \in W$ and $\theta_0(p) = p$,
(ii) $\theta_s \circ \theta_t(p) = \theta_{s+t}(p) = \theta_t \circ \theta_s(p)$

if all terms are defined. The infinitesimal generator X is defined by (3.2) just as before and is again $X = \partial/\partial x$. Finally we have orbits $t \to \theta_t(p)$, which are the lines $y = \text{constant}$ as before when $p = (x, y)$, $y \neq 0$, and for $p = (x, 0)$ the portion of the x-axis minus the origin which contains p. This curve is not defined for all values of t in the case of the orbit of a point on the x-axis. Exercise 10 gives an important example of a different type. A careful study of these examples will motivate the following complicated definition.

First let M be a C^∞ manifold and $W \subset R \times M$ be an *open* set which satisfies the following condition:

(3.10) *For every $p \in M$ there exist real numbers $\alpha(p) < 0 < \beta(p)$ such that $W \cap (R \times \{p\}) = \{(t, p) \mid \alpha(p) < t < \beta(p)\}$.*

We shall denote by $I(p)$ the interval $\alpha(p) < t < \beta(p)$ and by I_δ the interval defined by $|t| < \delta$. Condition (3.10) simply states that $W = \bigcup_{p \in M} I(p) \times \{p\}$. Note that W is connected. Using this notation and with W as above we make the following definition:

(3.11) Definition A *local one-parameter group action* or *flow* on a manifold M is a C^∞ map $\theta \colon W \to M$ which satisfies the following two conditions:

(i) $\theta_0(p) = p$ for all $p \in M$.
(ii) If $(s, p) \in W$, then $\alpha(\theta_s(p)) = \alpha(p) - s$, $\beta(\theta_s(p)) = \beta(p) - s$, and moreover for any t such that $\alpha(p) - s < t < \beta(p) - s$, $\theta_{t+s}(p)$ is defined and

$$\theta_t \circ \theta_s(p) = \theta_{t+s}(p).$$

It is easy to check that the Example 3.9 given above has these properties. This example also shows that if we are to have any prospect of obtaining a correspondence between one-parameter group actions and vector fields, we must abandon the requirement that W is all of $R \times M$, which we shall call a

global action. Since W is open and contains $(0, p)$ for each $p \in M$, it also contains $I_\delta \times U$, U a neighborhood of p, for sufficiently small $\delta > 0$. Therefore the definition of the vector field X (infinitesimal generator) associated with θ as given by (3.2) is valid in the case of local action also and associates a C^∞-vector field to each flow θ.

When R acts on M, as in the case of any group acting on M, for each t, $\theta_t \colon M \to M$ is a diffeomorphism with $\theta_t^{-1} = \theta_{-t}$. Something like this is also true for the local case of Definition 3.11, except that θ_t is not defined on all of M in general. Let $V_t \subset M$ be the domain of definition of θ_t, that is, $V_t = \{p \in M \mid (t, p) \in W\}$; then we have the following consequence of Definition 3.11:

(3.12) Theorem *V_t is an open set for every $t \in R$ and $\theta_t \colon V_t \to V_{-t}$ is a diffeomorphism with $\theta_t^{-1} = \theta_{-t}$.*

Proof Let $p_0 \in V_{t_0}$, so that $(t_0, p_0) \in W$. Since W is open, there is a $\delta > 0$ and a neighborhood V of p_0 such that $\{t \mid |t - t_0| < \delta\} \times V \subset W$. In particular, $\{t_0\} \times V \subset W$ so that $V \subset V_{t_0}$, so V_{t_0} is open. Next, by Definition 3.11(ii), if $p \in V_t$, then $\alpha(p) < t < \beta(p)$ and by (3.10) $t + (-t)$ lies in the same interval. It follows that $\theta_t(p) \in V_{-t}$ and $\theta_{-t} \circ \theta_t(p) = p$. Similarly, $\theta_{-t}(V_{-t}) \subset V_t$ and $\theta_t \circ \theta_{-t}(q) = q$ for any $q \in V_{-t}$. Combining these statements with the fact that θ_t, θ_{-t} are C^∞ on any open subsets of M on which they are defined completes the proof. ∎

(3.13) Remark For local one-parameter action we may show as in the global case that: $\theta_{t*}(X_p) = X_{\theta_t(p)}$ if $p \in V_t$. As before, $F(t) = \theta_t(p)$ defined for $\alpha(p) < t < \beta(p)$ is a C^∞-integral curve of X, which is an immersion of $I(p)$ in M provided that $X_p \neq 0$ and is a single point if $X_p = 0$. We shall continue to refer to these curves as *orbits* of the local one-parameter group, just as in the global case. It is a consequence of our definitions that these curves (and points) partition M into a union of mutually disjoint sets. The proofs are the same, essentially, as in the global case.

Finally, we wish to prove that in a neighborhood of any p for which $X_p \neq 0$, Example 3.9 is a prototype for every local (or global) one-parameter group action on a manifold.

(3.14) Theorem *Let $\theta \colon W \to M$ be as in Definition 3.11 and let X be the associated infinitesimal generator. If $p \in M$ such that $X_p \neq 0$, then there is a coordinate neighborhood V, ψ around p, a $\nu > 0$, and a corresponding neighborhood V' of p, $V' \subset V$, such that in local coordinates θ restricted to $I_\nu \times V'$ is given by*

$$(t, y^1, \ldots, y^n) \to (y^1 + t, y^2, \ldots, y^n).$$

In these coordinates $X = \psi_*^{-1}(\partial/\partial y^1)$ at every point of V'.

Proof We shall use the notation and formula for X_p developed in the discussion of (3.2). In W introduce coordinates U, φ around p and express θ in the local coordinates by $x \to h(t; x)$, where $x = (x^1, \ldots, x^n)$ and $h(t; x)$ stands for an n-tuple of functions satisfying: (i) $h(0; x) = x$, and (ii) $h(t; h(t'; x)) = h(t + t'; x)$. We will assume coordinates so chosen that $\varphi(p) = (0, \ldots, 0)$, that $\varphi(U) = C_\varepsilon^n(0)$, and that $X_p = \varphi_*^{-1}(\partial/\partial x^1) = E_{ip}$. Then the expression for X_p, $X_p = \sum \dot{h}^i(0; 0, \ldots, 0)E_{ip}$, implies that $\dot{h}^i(0; 0, \ldots, 0)$ is 1 for $i = 1$ and is 0 for $i > 1$.

Choose $\delta > 0$ small enough so that $V'' = \varphi^{-1}(C_\delta^n(0)) \subset U$ and $\theta(I_\delta \times V'') \subset U$. Then map the cube $C_\delta^n(0) \subset I_\delta \times \boldsymbol{R}^{n-1}$ into $\varphi(U)$ by a map F defined by

$$F\colon (y^1, \ldots, y^n) \to (h^1(y^1; 0, y^2, \ldots, y^n), \ldots, h^n(y^1; 0, y^2, \ldots, y^n)).$$

Since $X_p = \sum \dot{h}_i(0; 0)E_{ip} = E_{1p}$, $(\partial h^i/\partial y^1)_0 = \delta_1^i$; and from (ii), $y^i \equiv h^i(0; 0, y^2, \ldots, y^n)$, $(\partial h^i/\partial y^j)_0 = \delta_j^i$ for $j > 1$. Thus the Jacobian of F at $y = (0, \ldots, 0)$ is the identity matrix; hence there is a $\mu > 0$ with $\mu \le \delta$ such that F is a diffeomorphism of $C_\mu^n(0)$ onto an open set of $C_\varepsilon^n(0) = \varphi(U)$. Let $V = \varphi^{-1} \circ F(C_\mu^n(0))$ and $\psi = F^{-1} \circ \varphi$; this is a coordinate neighborhood of p with $V \subset U$.

The relations satisfied by $h^i(t, x)$, $i = 1, \ldots, n$, give

(i) $\psi(p) = F^{-1}(\varphi(p)) = F^{-1}(0, \ldots, 0)$

and for $(y^1, \ldots, y^n) \in C_v(0)$ and $|t| < v$ with $v = \mu/2$ they give

(ii) $h^i(t + y^1; 0, y^2, \ldots, y^n) = h^i(t, h(y^1; 0, y^2, \ldots, y^n))$, $i = 1, \ldots, n$.

Formula (ii) may be interpreted as follows: In the coordinate system (V, ψ), if $\psi(q) = (y^1, \ldots, y^n)$, then $\psi(\theta_t(q)) = (t + y^1, \ldots, y^n)$, provided only that $|t| < v$ and $q \in \psi^{-1}(C_v^n(0))$, so that all functions are defined. In other words, in the y-coordinates of V, ψ, the mapping θ_t is expressed by functions $\tilde{h}^i(t, y)$ defined on $I_v \times C_v^n(0)$ by

$$\tilde{h}^1(t, y^1, \ldots, y^n) = t + y^1,$$
$$\tilde{h}^i(t, y^1, \ldots, y^n) = y^i \qquad \text{for} \quad i > 1.$$

From these and the formula

$$\psi_*(X_q) = \sum \dot{h}^i(0, y)\frac{\partial}{\partial y^i} = \frac{\partial}{\partial y^1},$$

we have $X_q = \psi_*^{-1}(\partial/\partial y^1)$ on $V' = \psi^{-1}(C_v^n(0))$. ∎

Exercises

1. Let $F\colon M \to N$ be a C^∞ mapping whose rank is everywhere zero (that is, $F_* = 0$ at each $p \in M$). Show that F maps each component of M into a single point.

2. Let $F: M \to N$ be a C^∞ mapping and X and Y vector fields on M and N, respectively, which are F-related. Show that any integral curve of X is mapped by F into an integral curve of Y.
3. Give a proof without using local coordinates that the infinitesimal generator X of θ as defined in (3.2) is a C^∞-vector field on M. [*Hint:* Use Exercise 2.1].
4. Verify the statements of Remark 3.13.
5. Let R act on $M = \boldsymbol{R}^2$ according to the formulas

$$x \to x \cos t + y \sin t, \qquad y \to -x \sin t + y \cos t,$$

which give $\theta_t(x, y)$. Show that this is a globally defined group action of R on M and find X, the infinitesimal generator. What are the orbits?
6. Show that $(x, y) \to \theta(t; x, y)$, defined by

$$x \to xe^{2t}, \qquad y \to ye^{-3t}$$

defines a C^∞ action of R on $M = \boldsymbol{R}^2$. Determine the infinitesimal generator and show that it is θ-invariant.
7. Let $M = Gl(2, \boldsymbol{R})$ and define an action of R on M by the formula

$$\theta(t, A) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \cdot A, \qquad A \in Gl(2, \boldsymbol{R}),$$

with the dot denoting matrix multiplication. Find the infinitesimal generator.
8. Let $X = \sum_{i=1}^n f^i(x)(\partial/\partial x^i)$ be a C^∞-vector field on $\boldsymbol{R}^n$ which generates an action θ on $\boldsymbol{R}^n$. Suppose θ to be given on its domain W by $\theta(t, x) = (h^1(t, x), \ldots, h^n(t, x))$ and suppose $a = (a^1, \ldots, a^n) \in \boldsymbol{R}^n$. Then using Remark 3.13 show that $x^i = h^i(t, a)$, $i = 1, \ldots, n$, are solutions of the differential equations

$$\frac{dx^i}{dt} = f^i(x), \qquad i = 1, \ldots, n,$$

satisfying $x^i = a^i$, $i = 1, \ldots, n$, when $t = 0$. Verify this for Exercises 6–8.
9. Suppose that a C^∞ action of R on M, $M = \boldsymbol{R}^2$ has infinitesimal generator $X = x\, \partial/\partial x + y\, \partial/\partial y$ on M. Determine θ. [*Hint:* Try to find functions $h^1(t, x, y)$ and $h^2(t, x, y)$ by solving the system of ordinary differential equations $dx/dt = x$, $dy/dt = y$ with initial conditions $x(0) = a$, $y(0) = b$ as in Exercise 8.]
10. Let $X = x^2\, \partial/\partial x$ define a vector field on $M = \boldsymbol{R}$. Using the method of Exercise 8, find θ and its domain W.

4 The Existence Theorem for Ordinary Differential Equations

In this section we state a very basic theorem of analysis which we need in order to answer some of the questions raised in the previous section and which will be applied in essential ways throughout the remainder of the book.

(4.1) Theorem (Existence theorem for ordinary differential equations) *Let $U \subset \mathbf{R}^n$ be an open set and I_ε, $\varepsilon > 0$, denote the interval $-\varepsilon < t < \varepsilon$, $t \in \mathbf{R}$. Suppose $f^i(t, x^1, \ldots, x^n)$, $i = 1, \ldots, n$, to be functions of class C^r, $r \geq 1$, on $I_\varepsilon \times U$.*

Then for each $x \in U$ there exists $\delta > 0$ and a neighborhood V of x, $V \subset U$, such that:

(I) *For each $a = (a^1, \ldots, a^n) \in V$ there exists an n-tuple of C^{r+1} functions $x(t) = (x^1(t), \ldots, x^n(t))$, defined on I_δ and mapping I_δ into U, which satisfy the system of first-order differential equations*

$$(*) \qquad \frac{dx^i}{dt} = f^i(t, x), \qquad i = 1, \ldots, n,$$

and the initial conditions

$$(**) \qquad x^i(0) = a^i, \qquad i = 1, \ldots, n.$$

For each a the functions $x(t) = (x^1(t), \ldots, x^n(t))$ are uniquely determined in the sense that any other functions $\bar{x}(t), \ldots, \bar{x}^n(t)$ satisfying $(*)$ and $(**)$ must agree with $x(t)$ on an open interval around $t = 0$.

(II) *These functions being uniquely determined by $a = (a^1, \ldots, a^n)$ for every $a \in V$, we write them $x^i(t, a^1, \ldots, a^n)$, $i = 1, \ldots, n$, in which case they are of class C^r in all variables and thus determine a C^r map of $I_\delta \times V \to U$.*

A proof of (I), which uses the contracting mapping lemma is given in an Appendix to this section (see p.172). The proof of (II) is more difficult and may be found in the work of Hurewicz [1], Dieudonné [1], or Lang [1].

If the right-hand side of $(*)$ is independent of t, then the system of differential equations is called *autonomous*. Throughout the remainder of this chapter we shall deal only with autonomous systems. In this case it is possible to restate the hypotheses and conclusions of the fundamental existence theorem in coordinate-free form using the concepts of vector field and integral curve. This will allow us to derive various global theorems useful in both geometry and analysis from a purely local existence theorem about open subsets of $\mathbf{R}^n$.

We first reinterpret the existence theorem in the autonomous case, in which the functions f^i depend on $x = (x^1, \ldots, x^n)$ alone. For simplicity we

shall also assume hereafter that all data are C^∞. Define on $U \subset \boldsymbol{R}^n$ a C^∞-vector field X by

$$X = f^1(x)\frac{\partial}{\partial x^1} + \cdots + f^n(x)\frac{\partial}{\partial x^n}.$$

By Definition 3.8 an *integral curve* of X is a C^∞ mapping F of an open interval (α, β) of $\boldsymbol{R}$ into U such that $\dot{F}(t) = X_{F(t)}$ for all $\alpha < t < \beta$. If we write F in terms of its coordinate functions

$$F(t) = (x^1(t), \ldots, x^n(t)),$$

then the vector equation $\dot{F}(t) = X_{F(t)}$ is satisfied if and only if

$$\frac{dx^i}{dt} = f^i(x^1(t), \ldots, x^n(t)), \qquad i = 1, \ldots, n,$$

which states precisely that the functions $x(t) = (x^1(t), \ldots, x^n(t))$ are a solution of $(*)$. Given $x \in U$, (I) of Theorem 4.1 states that for each a in a neighborhood V of x there is a unique integral curve $F(t)$ (see Fig. IV.6), satisfying $F(0) = a$. $F(t)$ is defined at least for $-\delta < t < \delta$ where $\delta > 0$ is the same for every $a \in V$. If we use a notation for these integral curves through points of V which indicates dependence on the initial point a, say

$$F(t, a) = (x^1(t, a), \ldots, x^n(t, a)),$$

and use an overdot for differentiation with respect to t, these equations become

$$\dot{x}^i(t, a) = f^i(x(t, a)) \qquad \text{and} \qquad x^i(0, a) = a^i, \qquad i = 1, \ldots, n.$$

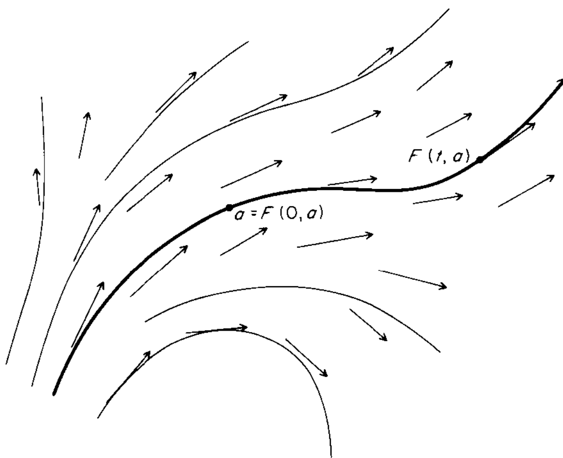


Figure IV.6
Integral curves of a vector field.

Part (II) of Theorem 4.1 states that these functions $x^i(t, a)$ are C^∞—in all variables—on $I_\delta \times V$, an open subset of $\boldsymbol{R} \times U$.

As an aid to intuition we may interpret the mapping $F: I_\delta \times V \to U$ as a "flow," that is, a motion within U of the points of V so that the point at position a at time $t = 0$ moves to $F(t, a)$ at time t (see Fig. IV.7). The path of a moving point is the integral curve, and its velocity at any of its positions is given by the vector X assigned to that point of U.

We now turn to a vector field X on an arbitrary manifold, considering first a purely local question.

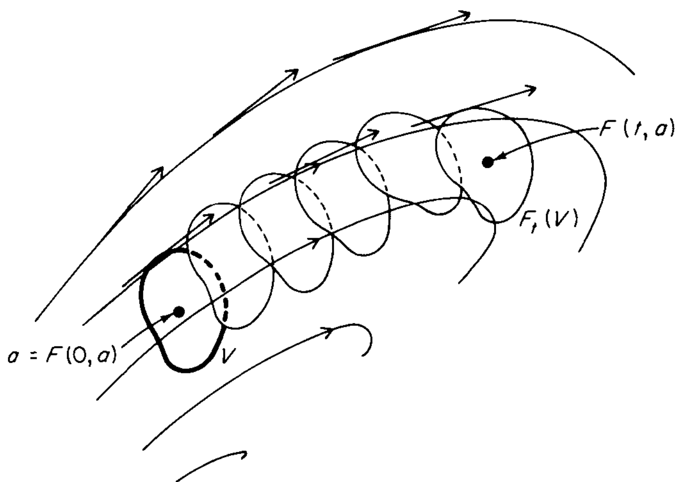


Figure IV.7

(4.2) Theorem *Let X be a C^∞-vector field on a manifold M. Then for each $p \in M$ there exists a neighborhood V and real number $\delta > 0$ such that there corresponds a C^∞ mapping*

$$\theta^V: I_\delta \times V \to M,$$

satisfying

$$(*) \qquad \dot{\theta}^V(t, q) = X_{\theta^V(t, q)}$$

and

$$(**) \qquad \theta^V(0, q) = q \qquad \text{for all} \quad q \in V.$$

If $F(t)$ is an integral curve of X with $F(0) = q \in V$, then $F(t) = \theta^V(t, q)$ for $|t| < \delta$. In particular, this mapping is unique in the sense that if V_1, δ_1 is another such pair for $p \in M$, then $\theta^V = \theta^{V_1}$ on the common part of their domains.

Proof This is basically a restatement of the existence theorem as follows. For $p \in M$ we choose a coordinate neighborhood U, φ and map X to the φ-related vector field $\tilde{X} = \varphi_*(X)$ on $\tilde{U} = \varphi(U) \subset \boldsymbol{R}^n$. After applying the local existence theorem to obtain $F: I_\delta \times \tilde{V} \to \tilde{U}$ defined by $F(t, a) = (x^1(t, a), \ldots, x^n(t, a))$ on a neighborhood $\tilde{V} \subset \tilde{U}$ of $\varphi(p)$, we set $V = \varphi^{-1}(\tilde{V})$ and define $\theta^V: I_\delta \times V \to U$ by $\theta^V(t, q) = \varphi^{-1}(F(t, \varphi(q)))$. Since φ and φ^{-1} are diffeomorphisms, we see at once that θ^V satisfies (*) and (**). The final assertion is a consequence of the uniqueness of solutions. ∎

Finally we consider the global aspects of the theory, that is, given a vector field X on M, what can be said that goes beyond the description of the situation in a neighborhood of a point. Our main purpose is to establish the relation between vector fields on M and local one-parameter groups acting on M (Theorem 4.6). The first result depends only on (I) of the existence theorem.

(4.3) Theorem *Let X be a C^∞-vector field on a manifold M and suppose $p \in M$. Then there is a uniquely determined open interval of R, $I(p) = \{\alpha(p) < t < \beta(p)\}$ containing $t = 0$ and having the properties:*

(1) *there exists a C^∞-integral curve $F(t)$ defined on $I(p)$ and such that $F(0) = p$;*

(2) *given any other integral curve $G(t)$ with $G(0) = p$, then the interval of definition of G is contained in $I(p)$ and $F(t) \equiv G(t)$ on this interval.*

Proof Let $F(t)$ and $G(t)$ be two integral curves such that $F(0) = p = G(0)$, and suppose I_F, I_G to be the open intervals on which they are defined, I^* the set on which they agree. I^* is not empty since it contains $t = 0$ and it is closed in $I_F \cap I_G$ since $F(t)$ and $G(t)$ are continuous. Suppose $s \in I^*$. Since $s \in I_F \cap I_G$, an open set, there is some interval $-\delta < t < \delta$ on which $\tilde{F}(t) = F(t + s)$ and $\tilde{G}(t) = G(t + s)$ are both defined. They are both integral curves satisfying the same initial condition: $\tilde{F}(0) = F(s) = G(s) = \tilde{G}(0)$. From the existence theorem they agree on some open interval $|t| < \delta$ around $t = 0$. Thus $F(t) = G(t)$ on an open set around s and I^* is open. It follows that $I^* = I_F \cap I_G$. Hence F and G define an integral curve on $I_F \cup I_G$. Thus, $I(p)$ is the union of the domains of all integral curves which pass through p at $t = 0$; the asserted properties are immediate. Note that it is quite possible for $\alpha(p) = -\infty$ and/or $\beta(p) = +\infty$. If both occur, then $I(p) = R$. ∎

We shall use the notation $F(t) = \theta(t, p)$ for the *unique maximal* integral curve $F(t)$ such that $F(0) = p$. When we wish to emphasize dependence on t, or p, we may write $\theta_p(t)$, or $\theta_t(p)$, for $\theta(t, p)$.

Now let the subset $W \subset R \times M$ be defined by

$$W = \{(t, p) \in R \times M \mid \alpha(p) < t < \beta(p)\}.$$

According to what has been shown thus far both W and θ are uniquely determined by X and W is the domain of $\theta: W \to M$. Moreover we have the following properties of θ and W:

(i) $\{0\} \times M \subset W$ and $\theta(0, p) = p$ for all $p \in M$.
(ii) For each (fixed) $p \in M$, let $\theta_p(t) = \theta(t, p)$. Then

$$\theta_p: I(p) \to M$$

is a C^∞-integral curve, that is, $\dot{\theta}_p(t) = X_{\theta_p(t)}$, and is maximal.
(iii) For each $p \in M$ there is a neighborhood V and a $\delta > 0$ such that $I_\delta \times V \subset W$ and θ is C^∞ on $I_\delta \times V$.

Using this notation and the same facts that we used in the proof of Theorem 4.3, we obtain the following addendum relating $I(p)$ and $I(q)$ for any two points p and q of the same integral curve.

(4.4) Corollary *Let $s \in I(p)$ and $q = \theta_p(s) = \theta(s, p)$ be the corresponding point of the integral curve determined by p. Then $\alpha(q) = \alpha(p) - s$ and $\beta(q) = \beta(p) - s$ so that*

$$I(q) = I(\theta_p(s)) = \{\alpha(p) - s < t < \beta(p) - s\}.$$

Thus $t \in I(q)$ if and only if $t + s \in I(p)$, and then we have

$$\theta(t, \theta(s, p)) = \theta(t + s, p).$$

Proof Suppose that $s \in I(p)$ and let $F(t) = \theta_p(s + t)$. Then $F(t)$ is defined on the open interval $\alpha(p) < s + t < \beta(p)$ and $F(0) = \theta_p(s) = q$ and is an integral curve. By maximality and by uniqueness we have $F(t) = \theta(t, \theta_p(s)) = \theta(t, q)$ so its domain must be $I(q) = \{\alpha(q) < t < \beta(q)\}$. ∎

We take note that what has been proved to this point answers any questions we might have about the existence of integral curves of a vector field X on a manifold. It does not describe completely the nature of W. We do that now and at the same time specify the relation between vector fields and one-parameter group action.

(4.5) Theorem *For any C^∞-vector field X the domain W of $\theta(t, p)$ is open in $R \times M$ and θ is a C^∞ map onto M.*

Proof We must show that for each $(t, p_0) \in W$ there is a neighborhood V of p_0 and $\delta > 0$ such that the open set $(t - \delta, t + \delta) \times V$ is in W and θ is C^∞ on it. This is already known to be the case for $(0, p_0)$. If the theorem fails,

then there must be some $(t_0, p_0) \in W$ such that for each $0 \le t < t_0$ there exists $(t - \delta, t + \delta) \times V$ with the above properties, but not for (t_0, p_0). [We have assumed, without loss of generality, that $t_0 > 0$).] We shall show by contradiction that there can be no (t_0, p_0).

Should there be such, then using Theorem 4.2, we find $\delta_0 > 0$ and a neighborhood V_0 of $q_0 = \theta(t_0, p_0)$ such that $I_{\delta_0} \times V_0 \subset W$ and θ is C^∞ on it. By continuity of $\theta(t, p_0)$ in t we may find $t_1 < t_0$ with both $|t_0 - t_1| < \frac{1}{3}\delta_0$ and $\theta(t_1, p_0) \in V_0$. Since $t_1 < t_0$, by our assumption on (t_0, p_0) there is a $\delta_1 > 0$ and a neighborhood V_1 of p_0 such that $(t_1 - \delta_1, t_1 + \delta_1) \times V_1 \subset W$ and such that θ is C^∞ on this open set. In particular, $\theta(t_1, p_0)$ is in V_0 and $\theta_{t_1}: V_1 \to M$ is defined and C^∞, so we may suppose by continuity (and restricting V_1 if necessary) that $\theta_{t_1}(V_1) \subset V_0$. We now have $\theta(s + t_1, q)$ defined and C^∞ on the open set $|s| < \delta_1$ and $q \in V_1$; and its values for $s = 0$ are in V_0. By Corollary 4.4 for $\alpha(\theta(t_1, q)) < s < \beta(\theta(t_1, q))$ the equation

$$\theta(s + t_1, q) = \theta(s, \theta(t_1, q))$$

is valid. Since $\theta(t_1, q)$ is in V_0, by the definition of δ_0 and V_0 the interval $I(\theta(t_1, q))$ contains all s for which $|s| < \delta_0$. Thus $\theta(s + t_1, q)$ is defined and C^∞ for $|s| < \delta_0$ and any $q \in V_1$. However, this is an open set containing (t_0, p_0) since $|t_0 - t_1| < \frac{1}{3}\delta_0$. This shows that our assumption on (t_0, p_0) leads to a contradiction. ∎

We now note that (4.3)–(4.5) tell us exactly that $\theta: W \to M$ defines a local one-parameter group action on M (Definition 3.11). It arises from X via the existence theorem and has X as infinitesimal generator. In fact it tells us somewhat more: Suppose $\theta_1 W \to M$ and $\theta_2: W_2 \to M$ are local one-parameter group actions. We will write $\theta_1 \cong \theta_2$ on $W_1 \cap W_2$. Collecting our results, we have the following, which is equivalent to the existence theorem in the autonomous case:

(4.6) Theorem *In the notation above, $\theta_1 \cong \theta_2$ if and only if they have the same infinitesimal generator X. Further, every C^∞ vector field X is the infinitesimal generator of a unique local one-parameter group action θ* on *M whose domain W is maximal among all $\tilde{\theta} \cong \theta$.*

(4.7) Remark It is important to note that now Theorems (3.12) and (3.14) hold for the $\theta(t, p)$ defined on W by applying the existence theorem. Thus, θ_t is a diffeomorphism of its domain $V_t = \{p \in M \mid t \in I(p)\}$, an open set, onto $V_{-t} = \{p \in M \mid -t \in I(p)\}$. Further, if $X_p \neq 0$, then there are local coordinates U, φ in which $X = E_1 = \varphi_*^{-1}(\partial/\partial x^1)$ on U.

(4.8) Remark A general nth order ordinary differential equation in the independent variable t and dependent variable x and its derivatives is given by a relation

$$F\left(t, x, \frac{dx}{dt}, \ldots, \frac{d^n x}{dt^n}\right) = 0.$$

We suppose that this is a function of class C^r defined on some neighborhood in $\boldsymbol{R}^{n+2}$ of the point $(0, a_0, a_1, a_2, \ldots, a_n)$ and that in a neighborhood U of this point we can write it in the form

$$\frac{d^n x}{dt^n} = G\left(t, x, \frac{dx}{dt}, \ldots, \frac{d^{n-1} x}{dt^{n-1}}\right).$$

(This can be done if the derivative of F with respect to its last variable is not zero at the point.)

Now let $x = x^1$, $dx/dt = x^2, \ldots, d^{n-1}x/dt^{n-1} = x^n$ and consider the first-order *system* of ordinary differential equations

$$(*) \qquad \frac{dx^1}{dt} = x^2, \quad \frac{dx^2}{dt} = x^3, \quad \ldots, \quad \frac{dx^n}{dt} = G(t, x^1, x^2, \ldots, x^{n-1}),$$

with initial conditions

$$(**) \qquad x^i(0) = a^i, \qquad i = 1, \ldots, n.$$

The original nth order equation has a solution $x(t)$ satisfying initial conditions (at $t = 0$):

$$x(0) = a^1, \quad \left(\frac{dx}{dt}\right)_0 = a^2, \quad \ldots, \quad \left(\frac{d^{n-1}x}{dt^{n-1}}\right)_0 = a^n$$

if and only if the first-order system $(*)$ has a solution satisfying $(**)$. Hence the existence theorem (Theorem 4.1) gives the existence and uniqueness of solutions of ordinary differential equations of nth order. This can be extended also to *systems* of ordinary differential equations of higher order than one. The conclusions of Theorem 4.1 concerning uniqueness of solutions and differentiability of dependence on initial conditions are also valid in this more general situation.

A second generalization is the case in which the functions f^i of system $(*)$ of Theorem 4.1 depend on parameters $z^1, \ldots, z^m$ so that the system becomes

$$\frac{dx^i}{dt} = f^i(t, x^1, \ldots, x^n, z^1, \ldots, z^m), \qquad i = 1, \ldots, n.$$

If we assume that the functions f^i are of class C^r in the z's also on some open set $U' \subset \mathbf{R}^m$, that is, f^i is a function of class C^r on $I_\varepsilon \times U \times U' \subset R \times \mathbf{R}^n \times \mathbf{R}^m$, then the solutions will depend on the z's as well as on the initial conditions:

$$x^i = x^i(t, a^1, \ldots, a^n, z^1, \ldots, z^m).$$

It is a further consequence of Theorem 4.1 that these functions are of class C^r in all variables on an open set $I_\varepsilon \times V \times V' \subset R \times \mathbf{R}^n \times \mathbf{R}^m$. This is very easily proved by introducing new equations of the form $dz^j/dt = 0, j = 1, \ldots, m$, so that we are dealing with a system of $n + m$ ordinary equations to which we apply Theorem 4.1.

An application will be made of this idea in the following case. Choose a basis $E_1, \ldots, E_n$ of the tangent space at the identity e of a Lie group G and let $X_g(z^1, \ldots, z^n)$ denote the value at $g \in G$ of the uniquely determined left-invariant vector field X whose value X_e at e has components $z^1, \ldots, z^n$, that is, $X_e = \sum_{i=1}^n z^i E_i$. With the choice of basis fixed, the left-invariant vector fields on G are thus parametrized by $\mathbf{R}^n$. The dependence on g and on the parameters is C^∞ so that the solutions of the system of equations corresponding to each of the vector fields $X(z^1, \ldots, z^n)$ is C^∞ in all variables. Thus we have $\theta(t; g; z^1, \ldots, z^n)$, which gives a C^∞ mapping $\theta\colon R \times G \times \mathbf{R}^n \to G$ and, for g, z fixed, determines the integral curve through g.

Exercises

1. Consider a system of n ordinary differential equations of second order in n unknown functions:

$$\frac{d^2x^k}{dt^2} = f^k\left(t, x^1, \ldots, x^n, \frac{dx^1}{dt}, \ldots, \frac{dx^n}{dt}\right), \qquad k = 1, \ldots, n.$$

   State as precisely as you can an existence theorem for solutions and derive it from Theorem 4..1.
2. Give a detailed statement and proof of the two generalizations indicated in Remark 4.8.
3. Let $M = \mathbf{R}^2$, the xy-plane, and $X = y(\partial/\partial x) - x(\partial/\partial y)$. Find the domain W and the one-parameter group $\theta\colon W \to M$.
4. Let X and Y be vector fields on manifolds M and N, respectively, and $F\colon M \to N$ a C^∞ mapping. Show that X and Y are F-related if and only if the local one-parameter groups θ and σ generated by X and Y satisfy $F \circ \theta_t(p) = \sigma_t \circ F(p)$ for all (t, p) for which both sides are defined.
5. Give a precise meaning to the following statement and then prove it: If the vector field X on M generates the local one-parameter group θ acting on M, X is invariant under the action.

6. Show that the orbits of a local one-parameter group may be defined in terms of an equivalence relation, just as in the case of a group G acting on a manifold. Show by example that the orbit space may fail to be Hausdorff.
7. Show that the general (nonautonomous) system on $\boldsymbol{R}^n$ of Theorem 4.1 can be reduced to the autonomous case on $\boldsymbol{R}^{n+1}$ by letting $t = x^{n+1}$ and adding an equation $dx^{n+1}/dt = 1$ with $x^{n+1}(0) = 0$.
8. Derive Theorem 4.1 from Theorem 4.6.

5 Some Examples of One-Parameter Groups Acting on a Manifold

We shall now consider a local one-parameter group θ with (maximal) domain W and infinitesimal generator X acting on a manifold M. For $p \in M$, we continue to denote by $I(p)$ the set $\alpha(p) < t < \beta(p)$ of all real numbers t such that (t, p) is in W. The integral curve of X through p is given by $\theta_p\colon I(p) \to M$, $\theta_p(t) = \theta(t, p)$. If $X_p = 0$, the curve is a single point p; otherwise θ_p is an immersion as was shown earlier. In this latter case we consider now the nature of the integral curves on M.

(5.1) Lemma *Suppose that $\beta(p) < \infty$ and that $\{t_n\} \subset I(p)$ is an increasing sequence converging to $\beta(p)$. Then $\{\theta(t_n, p)\}$ cannot lie in any compact set. In particular, the sequence $\{\theta(t_n, p)\}$ cannot approach a limit on M. A similar statement holds for a decreasing sequence approaching $\alpha(p)$ if $\alpha(p)$ is finite.*

Proof Let K be a compact subset of M and X a C^∞-vector field on M. By the existence theorem to each $q \in M$ corresponds a $\delta > 0$ and a neighborhood V of q such that θ is defined on $I_\delta \times V$. A finite number of such neighborhoods cover K and we let δ_0 be the minimum δ for these neighborhoods. Then for each $q \in K$, $\theta(t, q)$ is defined for $|t| < \delta_0$. Suppose $\{\theta(t_n, p)\} \subset K$ and that N is so large that $\beta(p) - t_N < \frac{1}{3}\delta_0$. Then, by Corollary 4.4, $\theta(t, \theta(t_N, p))$ is defined only for $t < \beta(p) - t_N < \delta_0/3$. But this is a contradiction since $\theta(t_N, p) \in K$, so $\theta(t, \theta(t_N, p))$ is defined at least for $-\delta_0 < t < \delta_0$. This proves the first statement. The second is an immediate consequence for if $\lim_{n\to\infty} \theta(t_n, p) = q$, then there is a neighborhood of q whose closure K is compact and contains all but a finite number of terms of the sequence $\{\theta(t_n, p)\}$. We discard the terms not in K and obtain the same contradiction. Obviously the same arguments apply to decreasing sequences approaching $\alpha(p)$—if $\alpha(p)$ is finite. ∎

(5.2) Corollary *If $I(p)$ is a bounded interval, then the integral curve is a closed subset of M.*

(5.3) Corollary *If $X_p = 0$, then $I(p) = \boldsymbol{R}$. If $X = 0$ outside a compact subset of M, then $W = \boldsymbol{R} \times M$.*

We leave the proofs of these two corollaries to the exercises.

(5.4) Remark A point p of M at which $X_p = 0$ is called a *singular* point of the vector field and any other point is referred to as *regular*. We have seen (Theorem 3.14) that in the neighborhood of a regular point the integral curves are—to within diffeomorphism—the family of parallel lines $x^2 = c^2, \ldots, x^n = c^n$ in $\boldsymbol{R}^n$. On the other hand the pattern of integral curves at an isolated singularity can take many forms, even in the two-dimensional case, and has been extensively studied. At least in the two-dimensional case singularities can be visualized in terms of the integral curves of the vector field X near p. Some possibilities are shown in Fig. IV.8a–d. The cases (a) and (b) correspond to the field

$$X = \operatorname{grad} f = \frac{\partial f}{\partial x}\frac{\partial}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial}{\partial y}$$

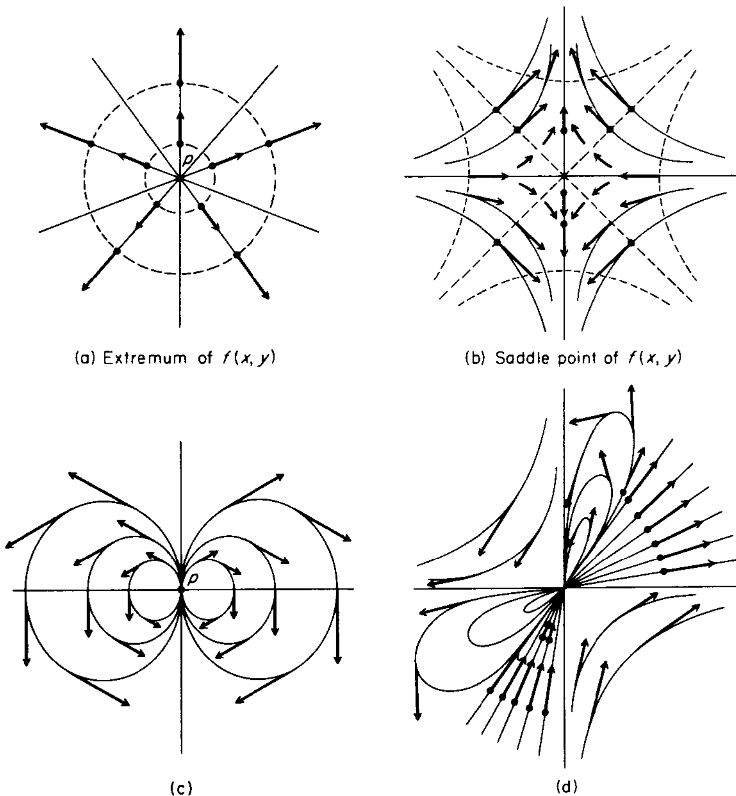


Figure IV.8

at a point $p = (x_0, y_0)$ which is a simple extremum or saddle point of the function $f(x, y)$. The level curves $f(x, y) =$ constant are the dotted lines orthogonal to the integral curves. If p is a singularity of a general vector field, the pattern can be more complicated; possibilities are shown in (c) and (d).

Interesting relations between the topological nature of the surface and the possible types of singularities possessed by a vector field on it were discovered by Poincaré, Hopf, and others (see Milnor [2] and Arnold [1]). These relations depend on the concept of the index i_p of a vector field X at an isolated singularity p. In the two-dimensional case this is an integer defined as follows: Let C be a simple closed curve containing p in its interior—and no other singular point. As we traverse C counterclockwise, the vector field X_q, $q \in C$, rotates, i.e., its angle θ with a fixed direction changes by an integral multiple $2\pi i_p$ of 2π. (In Fig. IV.8a,b,c, the index of X is $+1$, -1, 2, respectively.) The results alluded to show that i_p does not depend on the C chosen and amazingly that if X on a compact orientable surface M has only isolated singularities, then the sum of the indices is the Euler characteristic χ. In particular, it is independent of X. A consequence of these relations is the fact already mentioned that a vector field on S^2—in fact on any closed orientable surface except T^2—must have at least one singular point.

Another important question about a vector field X on M is whether or not it has closed integral curves—diffeomorphic to the circle S^1 (see Exercise 3). This can be of importance, for instance, in applications to dynamics. In these applications one considers the points of a manifold as corresponding to, or parametrizing, the states of a dynamical system. For example, if the system consists of the earth, sun, and moon, then in a fixed coordinate system the positions of the three objects can be characterized by nine numbers (three sets of coordinates) and their velocities, or momenta, by nine more (the components of three vectors). Thus each state or configuration corresponds to a point on a manifold M of dimension 18. The laws of motion can be expressed as a system of ordinary differential equations or vector field X on M, and the integral curves correspond to the motions beginning from various initial states. A closed integral curve corresponds to a periodic motion, like that of the planets. This approach to mechanics was extensively studied by Poincaré and Birkhoff, and is still an active area of research (see Smale [2]). It has led to many interesting questions about vector fields and curves on manifolds. For example, it was very recently shown by Schweitzer [1], that there exist everywhere regular vector fields on S^3 without any closed integral curves—contrary to a long standing conjecture. Classical mechanics in the framework of manifold theory is very clearly set forth by Godbillon [1]. Other excellent books on differential equations and dynamical systems are Hirsch and Smale [1], Arnold [2], and Abraham and Marsden [1].

(5.5) Definition A vector field X on M is said to be *complete* if it generates a (global) action of R on M, that is, if $W = R \times M$.

This is clearly the most desirable case and we find it very convenient to have sufficient conditions for completeness. One of them is an immediate corollary of Lemma 5.1.

(5.6) Corollary *If M is a compact manifold, then every vector field X on M is complete.*

To see that this is so we take $K = M$ in the lemma and note that in this case $\alpha(p) = -\infty$ and $\beta(p) = +\infty$, that is, $I(p) = R$, for every $p \in M$.

This gives one important case in which we may be sure that a vector field is complete. A second case, which we will study in some detail, is a left-invariant vector field on a Lie group, as is shown by the corollary to the theorem which follows.

(5.7) Theorem *Let X be a C^∞-vector field on a manifold M and $F: M \to M$ a diffeomorphism. Let $\theta(t, p)$ denote the C^∞ map $\theta: W \to M$ defined by X. Then X is invariant under F if and only if $F(\theta(t, p)) = \theta(t, F(p))$ whenever both sides are defined.*

Proof Suppose that X is invariant under F. If $\theta_p: I(p) \to M$ is the integral curve of X with $\theta_p(0) = p$, then the diffeomorphism F takes it to an integral curve $F(\theta_p(t))$ of the vector field $F_*(X)$. Since $F_*(X) = X$ and $F(\theta_p(0)) = F(p)$, from uniqueness of integral curves we conclude that $F(\theta_p(t)) = \theta(t, F(p))$. This proves the "only if" part of the theorem.

Now suppose that $F(\theta(t, p)) = \theta(t, F(p))$ and prove that $F_*(X_p) = X_{F(p)}$. This could be done directly from expression (3.2) for the infinitesimal generator X, but we shall proceed in a slightly different way. Let $\theta_p(t) = \theta(t, p)$ and let d/dt be the natural basis of $T_0(R)$, the tangent space to R at $t = 0$. Then, by definition, $X_p = \dot{\theta}_p(0) = \theta_{p*}(d/dt)$ and applying the isomorphism $F_*: T_p(M) \to T_{F(p)}(M)$ to this definition we have

$$F_*(X_p) = F_* \circ \theta_{p*}(d/dt) = (F \circ \theta_p)_*(d/dt) = \theta_{F(p)*}(d/dt) = X_{F(p)}.$$

The second equality is the chain rule for the composition of mappings applied to $\theta_p: R \to M$ and $F: M \to M$. The third equality uses the hypothesis that $F \circ \theta_p(t) = \theta_{F(p)}(t)$. ∎

We remark that in the notation of Section 3 this theorem could be stated: $F_*(X) = X$ *if and only if* $\theta_t \circ F = F \circ \theta_t$ on V_t.

(5.8) Corollary *A left-invariant vector field on a Lie group G is complete.*

Proof Let X be such a vector field. There is a neighborhood V of e and a $\delta > 0$ such that $\theta(t, g)$ is defined on $I_\delta \times V$. For $h \in G$, let L_h denote the left translation by h. If we apply Theorem 5.7 with $F = L_h$, then $\theta(t, L_h g) = L_h \theta(t, g)$, which shows that θ is defined on $I_\delta \times L_h(V)$, a neighborhood of $(0, h)$ in $R \times G$. It follows that for every $h \in G$ there is a neighborhood $U = L_h(V)$ such that $I_\delta \times U \subset W$, the domain of θ with the same $\delta > 0$ that we first obtained for V, that is, δ is fixed and independent of h. By the same argument as in the compact case we obtain a contradiction if we assume for any $g \in M$ that either $\alpha(g)$ or $\beta(g)$ is finite. Therefore $W = R \times M$ and X is complete. ∎

We shall now carry our analysis of the Lie group situation somewhat further and in this way will obtain a number of examples of actions of R on manifolds.

(5.9) Definition Let R be the additive group of real numbers, considered as a Lie group, and let G be an arbitrary Lie group. A *one-parameter subgroup* H of G is the homomorphic image $H = F(R)$ of a homomorphism $F: R \to G$. It is called trivial if $H = \{e\}$.

We give here several simple examples of one-parameter subgroups. In the next section it will be shown how all such subgroups may be determined for linear Lie groups, that is, subgroups of $Gl(n, \mathbf{R})$. Since we are interested in the action of R on manifolds, we recall at this point a comment and examples of Section III.7. Namely, let G be a Lie group which acts on a manifold M by $\tilde{\theta}: G \times M \to M$ and let $F: R \to G$ be a homomorphism. Then $\theta: R \times M \to M$ defined by $\theta(t, p) = \tilde{\theta}(F(t), p)$ defines an action of R on M. Now applying Sections 3 and 4 we have an associated infinitesimal generator X, integral curves as orbits of the action, and so on. Since the same G may act on different manifolds, or in different ways on the same manifold, a fixed one-parameter subgroup of G will give many examples of a one-parameter group of transformations of a manifold.

(5.10) Example Let G be the group $Gl(3, \mathbf{R})$. We consider two one-parameter subgroups, that is, two homomorphisms F_1, F_2 of R into G, defined as follows ($a, b, c \in \mathbf{R}$ are constants):

$$F_1(t) = \begin{pmatrix} e^{at} & 0 & 0 \\ 0 & e^{at} & 0 \\ 0 & 0 & e^{at} \end{pmatrix} \quad \text{and} \quad F_2(t) = \begin{pmatrix} 1 & at & bt + \frac{1}{2}act^2 \\ 0 & 1 & ct \\ 0 & 0 & 1 \end{pmatrix}.$$

It is left as an exercise to check that these are homomorphisms. Now $Gl(3, \mathbf{R})$ acts naturally on $\mathbf{R}^3$ (Example III.7.4) and hence each F_i defines an

action on $\boldsymbol{R}^3$. In the case of F_1 we have $\theta(t, x^1, x^2, x^3) = (e^{at}x^1, e^{at}x^2, e^{at}x^3)$. Therefore the infinitesimal generator X is given at $x \in \boldsymbol{R}^3$ by

$$X_x = \dot{\theta}(0, x) = ax^1 \frac{\partial}{\partial x^1} + ax^2 \frac{\partial}{\partial x^2} + ax^3 \frac{\partial}{\partial x^3},$$

and the integral curves, or orbits, are the lines through the origin (see Fig. IV.9).

The group $Gl(n, \boldsymbol{R})$ also acts on $P^{n-1}(\boldsymbol{R})$, since it preserves the equivalence relation (proportionality) of n-tuples which defines it. Therefore $Gl(3, \boldsymbol{R})$ acts on two-dimensional projective space $P^2(\boldsymbol{R})$. In this case F_1 defines a trivial action $\theta(t, p) \equiv p$.

(5.11) Example Let G be the Lie group $SO(3)$ of orthogonal matrices with determinant $+1$. Define $F\colon R \to SO(3)$ and thus a one-parameter subgroup by

$$F(t) = \begin{pmatrix} \cos at & \sin at & 0 \\ -\sin at & \cos at & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Again, it is easily checked directly that this is in fact a homomorphism. Thus $SO(3)$ acts on the unit sphere S^2 in a standard manner which we previously discussed (Section III.7 and Exercises). The action is just the usual rotation of the sphere, and F defines a one-parameter group of rotations holding the x^3 axis fixed:

$$\theta(t, x^1, x^2, x^3) = (x^1 \cos at + x^2 \sin at, -x^1 \sin at + x^2 \cos at, x^3).$$

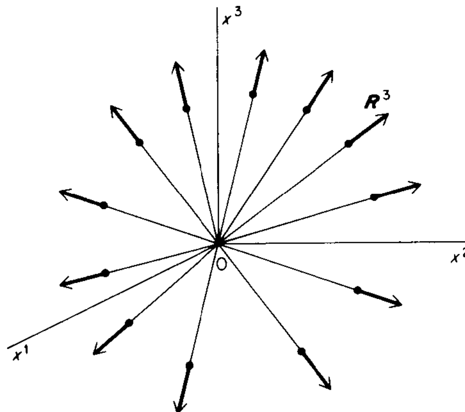


Figure IV.9

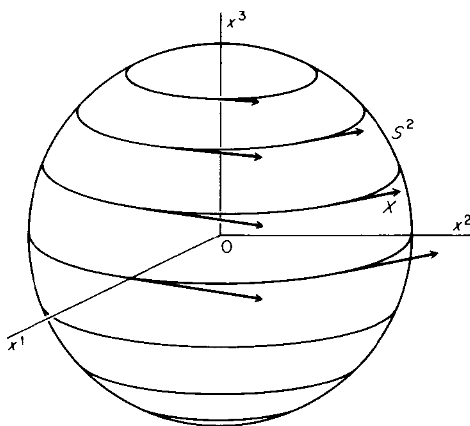


Figure IV.10

The orbits are the lines of latitude and the generator X is tangent to them and orthogonal to the x^3-axis. $X = 0$ at the north and south poles $(0, 0, \pm 1)$. (See Fig. IV.10.)

(5.12) Example We recall also that a Lie group G acts on itself (on the right) by right translations. Thus if we are given a homomorphism $F: R \to G$, we may define an action θ of R on $M = G$ by $\theta(t, g) = R_{F(t)}(g) = gF(t)$. We have used R_a to denote right translation: $R_a(g) = ga$. As previously noted in Section III.7, this is a composition of C^∞ maps, F, and right translation. It is an action since F is a homomorphism and multiplication is associative:

(i) $\theta(0, g) = gF(0) = g,$
(ii) $\theta(t + s, g) = gF(t + s) = g(F(t)F(s))$
$\qquad = (gF(t))F(s) = \theta(t, \theta(s, g)).$

Thus the examples above furnish further examples of one-parameter group action, namely, on $M = Gl(3, \boldsymbol{R})$ and $M = O(3)$, respectively.

Recalling that a left-invariant vector field on G is uniquely determined by its value at the identity e, we may use these ideas to characterize one-parameter subgroups of a Lie group.

(5.13) Theorem *Let $F: R \to G$ be a one-parameter subgroup of the Lie group G and X the left-invariant vector field on G defined by $X_e = \dot{F}(0)$. Then $\theta(t, g) = R_{F(t)}(g)$ defines an action $\theta: R \times G \to G$ of R on G (as a manifold) having X as infinitesimal generator. Conversely, let X be a left-invariant vector field and $\theta: R \times G \to G$ the corresponding action. Then $F(t) = \theta(t, e)$ is a one-parameter subgroup of G and $\theta(t, g) = R_{F(t)}(g)$.*

Proof Given the C^∞ homomorphism $F: R \to G$, then $\theta: R \times G \to G$, defined by $\theta(t, g) = R_{F(t)}(g) = gF(t)$ is, as we have just seen, an action of R on G. If $a \in G$, then $L_a\theta(t, g) = a(gF(t)) = (ag)F(t) = \theta(t, L_a(g))$. By Theorem 5.7 it follows that the generator X of θ is L_a-invariant, for any $a \in G$. However, $\theta(t, e) = F(t)$, and so $X_e = \dot{\theta}(0, e) = \dot{F}(0)$, which proves the first half of the theorem.

For the converse X, being left-invariant, is both C^∞ and complete and it generates an action θ of R on G. By Theorem 5.7 for any left translation L_h we have $L_h\theta(t, g) = \theta(t, L_h(g))$ or equivalently, $h\theta(t, g) = \theta(t, hg)$. Let $F(t) = \theta(t, e)$ and $h = F(s)$. Then this relation implies

$$F(s)F(t) = F(s)\theta(t, e) = \theta(t, \theta(s, e)) = \theta(t + s, e) = F(s + t).$$

Thus $t \to F(t)$ is a C^∞ homomorphism. But $\dot{F}(0) = \dot{\theta}(0, e) = X_e$ and since X is left-invariant, we see by uniqueness of the action generated by X that $\theta(t, g) = R_{F(t)}(g)$, the action defined just previously. ▮

(5.14) Corollary *There is a one-to-one correspondence between the elements of $T_e(G)$ and one-parameter subgroups of G. For $Z \in T_e(G)$ let $t \to F(t, Z)$ denote the (unique) corresponding one-parameter subgroup. Then $F: R \times T_e(G) \to G$ is C^∞ and satisfies $F(t, sZ) = F(st, Z)$.*

Proof According to Theorem 5.13, each $Z \in T_e(G)$ determines a unique homomorphism $t \to F(t, Z)$ of R into G such that $\dot{F}(0, Z) = Z$. By our extension of the existence theorem at the end of Remark 4.7 we see that F is C^∞ simultaneously in t and Z [identifying $T_e(G)$ with $\boldsymbol{R}^n$ by some choice of basis]. Using the rule for change of parameter in a curve on a manifold, we have

$$\left[\frac{d}{dt} F(ts, Z)\right]_{t=0} = s\left[\frac{d}{dt} F(t, Z)\right]_{t=0} = sZ.$$

On the other hand $t \to F(ts, Z)$ is a homomorphism. Therefore, by uniqueness, $F(st, Z) = F(t, sZ)$. ▮

Exercises

1. Prove Corollary 5.2.
2. Prove Corollary 5.3.
3. Let X be a vector field on M and let $F: I(p) \to M$ be the integral curve determined by $F(0) = p$. Suppose for some real number $c > 0$, $F(c) = F(0)$. Show that this implies $I(p) = R$ and $F(t) = F(t + c)$ for all $t \in R$. If $X_p \neq 0$, then prove that there is a diffeomorphism $G: S^1 \to M$ and a number c_0, $0 < c_0 \leq c$, such that $F = G \circ \pi$, with $\pi: R \to S^1 = \{z \in \boldsymbol{C} \mid |z| = 1\}$ denoting the mapping $\pi(t) = e^{2\pi i(t/c_0)}$.
4. Given $p \in M$, show that if $I(p)$ is bounded for a C^∞-vector field X on M, then $t \to \theta(t, p)$ is an imbedding of $I(p)$ in M.

5. Given $p \in M$ and a C^∞-vector field X on M, let $\{t_n\}$ be a monotone increasing (decreasing) sequence of $I(p)$ which has no limit on $I(p)$. Show that if $\lim_{n\to\infty} \theta(t_n, p)$ exists, then $\alpha(p) = -\infty$ ($\beta(p) = +\infty$, respectively). Let $L^+(p)$ [or $L^-(p)$] denote the collection of all such limit points for increasing (decreasing) sequences. Show that $L^\pm(\theta(t, p)) = L^\pm(p)$ for every $t \in I(p)$ and that $L^\pm(p)$ is a closed set and a union of integral curves.
6. Show that a one-parameter subgroup H of a Lie group G which is not trivial is either an isomorphic image of S^1 or R; and in the latter case is a submanifold whose closure is an (algebraic) Abelian subgroup, possibly larger than H. Give examples of each case.

6 One-Parameter Subgroups of Lie Groups

We have seen that one-parameter subgroups of a Lie group G are in one-to-one correspondence with the elements of $T_e(G)$. We shall use this to help determine all one-parameter subgroups of various matrix groups. We first consider $G = Gl(n, \boldsymbol{R})$. The matrix entries x_{ij}, $1 \le i, j \le n$, for any $X = (x_{ij}) \in Gl(n, \boldsymbol{R})$ are coordinates on a single neighborhood covering the group, which is an open subset of $\mathcal{M}_n(\boldsymbol{R})$, the $n \times n$ matrices over $\boldsymbol{R}$. Therefore $\partial/\partial x_{ij}$, $1 \le i, j \le n$, is a field of frames on G and, relative to these frames as a basis at e, there is an isomorphism of $\mathcal{M}_n(\boldsymbol{R})$ as a vector space onto $T_e(G)$ given by $A = (a_{ij}) \to \sum_{i,j} a_{ij}(\partial/\partial x_{ij})_e$. [When $G = Gl(n, \boldsymbol{R})$, e is the $n \times n$ identity matrix I.]

(6.1) Definition The *exponential* e^X *of a matrix* $X \in \mathcal{M}_n(\boldsymbol{R})$ is defined to be the matrix given by

$$(*) \qquad e^X = I + X + \frac{1}{2!}X^2 + \frac{1}{3!}X^3 + \cdots$$

if the series converges.

(6.2) Theorem *Series* $(*)$ *converges absolutely for all* $X \in \mathcal{M}_n(\boldsymbol{R})$ *and uniformly on compact subsets. The mapping* $\mathcal{M}_n(\boldsymbol{R}) \to \mathcal{M}_n(\boldsymbol{R})$ *defined by* $X \to e^X$ *is* C^∞ *and has nonsingular Jacobian at* $X = 0$. *Its image lies in* $Gl(n, \boldsymbol{R})$. *If* $A, B \in \mathcal{M}^n(\boldsymbol{R})$ *such that* $AB = BA$, *then* $e^{A+B} = e^A e^B$.

Proof If we denote by $x_{ij}^{(k)}$ the entries of the matrix X^k with $X^1 = X = (x_{ij})$ and $X^0 = I = (\delta_{ij})$, and we let $\rho = \sup_{1 \le i, j \le n} |x_{ij}|$, then by induction on k we have the inequality

$$|x_{ij}^{(k)}| \le (n\rho)^k.$$

This is true for $k = 0$, and if it holds for k, then

$$|x_{ij}^{(k+1)}| = \left|\sum_l x_{il}^{(k)} x_{lj}\right| \le n(n\rho)^k\rho = (n\rho)^{k+1}.$$

From this it follows that the sequence e^X converges absolutely for every X and that it converges uniformly on every compact subset of $\mathcal{M}_n(\boldsymbol{R})$; indeed each compact set is contained in a set $K_\rho = \{X \,|\, |x_{ij}| \le \rho\}$. Using absolute convergence, each entry of e^X may be written as a power series in any of the x_{ij} converging on all of $\boldsymbol{R}$ (or even $\boldsymbol{C}$). Hence $X \to e^X$ is analytic in each x_{ij} and is therefore an analytic mapping (by Hartog's theorem).

If we denote by $f_{ij}(X)$ the coordinate functions of the mapping, then the terms of degree less than 2 in the variables x_{ij} are

$$f_{ij}(X) = \delta_{ij} + x_{ij}, \qquad 1 \le i, j \le n;$$

hence the Jacobian matrix at $X = 0$ reduces to the $n^2 \times n^2$ identity matrix.

Finally, using the fact that the convergence is absolute, so that we may rearrange terms, and an analog of Cauchy's theorem for multiplication of series, when $AB = BA$ we obtain the equality

$$\left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k\right)\left(\sum_{l=0}^{\infty} \frac{1}{l!} B^l\right) = \sum_{m=0}^{\infty} \sum_{p=0}^{m} \frac{1}{(m-p)!} A^{m-p} \frac{1}{p!} B^p = \sum_m \frac{1}{m!}(A + B)^m.$$

From this we may deduce $e^A e^B = e^{A+B}$ (see Exercise 1). In particular, this implies $e^A e^{-A} = e^0 = I$ so that e^A is nonsingular, that is, $e^A \in Gl(n, \mathbf{R})$ for any $a \in \mathcal{M}_n(\boldsymbol{R})$. This completes the proof of the theorem. ∎

(6.3) Corollary *$F(t) = e^{tA}$ is the one-parameter subgroup of $Gl(n, \mathbf{R})$ whose corresponding left-invariant vector field has the value $\sum_{i,j} a_{ij}(\partial/\partial x_{ij})_e$. All one-parameter subgroups of $Gl(n, \mathbf{R})$ are of this form. Further, $\dot{F}(0) = A$.*

Proof The corollary is an immediate consequence of the theorem. For every $t \in \boldsymbol{R}$, $t_1 A$ and $t_2 A$ commute, thus $e^{(t_1+t_2)A} = e^{t_1 A} e^{t_2 A}$ and $t \to e^{tA}$ is a group homomorphism; it is C^∞ since it is a restriction of a C^∞-map on $\mathcal{M}_n(\boldsymbol{R})$ to the submanifold $\{tA \,|\, t \in \boldsymbol{R}\}$. Finally, writing $x_{ij}(t)$ for the ijth entry of e^{tA} and letting $A = (a_{ij})$, we have

$$x_{ij}(t) = \delta_{ij} + ta_{ij} + O(t^2)$$

so that $\dot{x}_{ij}(0) = a_{ij}$, $1 \le i, j \le n$ or, equivalently $(de^{tA}/dt)_{t=0} = A$. This proves the corollary. ∎

(6.4) Example

$$A = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix} \in \mathscr{M}_3(R)$$

implies

$$e^{tA} = I + tA + \tfrac{1}{2}t^2A^2 + \cdots.$$

However, $A^k = 0$ if $k > 2$ so that we obtain once again Example 5.10:

$$e^{tA} = \begin{pmatrix} 1 & ta & tb + \frac{1}{2}t^2ac \\ 0 & 1 & tc \\ 0 & 0 & 1 \end{pmatrix}.$$

By virtue of the following theorem, we can use the mapping $X \to e^X$ to determine the one-parameter subgroups of other matrix groups, for example, $O(n)$, $Sl(n, \boldsymbol{R})$ (see Exercise 6), and so on.

(6.5) Theorem *If H is a Lie subgroup of G, then the one-parameter subgroups of H are exactly those one-parameter subgroups $t \to F(t)$ of G such that $\dot{F}(0) \in T_e(H)$ considered as a subspace of $T_e(G)$.*

Proof Let $F: R \to H$ be any one-parameter subgroup of H. Since $H \subset G$ and the inclusion is an immersion, in particular is C^∞, the map F followed by inclusion is a one-parameter subgroup of G. Its tangent vector at any point is tangent to H. In particular, $\dot{F}(0) \in T_e(H)$ a subspace of $T_e(G)$. Conversely, if $F: R \to G$ is a one-parameter subgroup such that $\dot{F}(0) \in T_e(H)$, then $\dot{F}(0)$ determines a one-parameter subgroup of H, $F_1: R \to H$, with $\dot{F}_1(0) = \dot{F}(0)$. As we have just seen, F_1 can be considered a one-parameter subgroup of G, but since F and F_1 have the same tangent vector at e, they must agree. Therefore the correspondence is one-to-one as claimed, which completes the proof. ∎

Suppose that $G = Gl(n, \boldsymbol{R})$ in the discussion above, then we have the following application.

(6.6) Corollary *The one-parameter subgroups of a subgroup $H \subset Gl(n, \boldsymbol{R})$ are all of the form $F(t) = e^{tA}$, where $A = (a_{ij})$ are the components of the vector $\dot{F}(0) = \sum_{i,j} a_{ij}(\partial/\partial x_{ij})_e \in T_e(G)$ which is tangent to H at e, i.e., is in $T_e(H) \subset T_e(G)$.*

This is an immediate consequence of the theorem and the fact that every one-parameter subgroup of $G = Gl(n, \boldsymbol{R})$ is of the form $F(t) = e^{tA}$.

(6.7) Example Let $H = O(n)$, $G = Gl(n, \mathbf{R})$, and determine the one-parameter subgroups of H. If $e^{tA} \in H$ for all t, then $(e^{tA})(e^{tA})' = I$, where the prime indicates the transpose. It is an immediate consequence of Definition 6.1 that $(e^{tA})' = e^{tA'}$; and, by Theorem 6.2, $(e^{tA})^{-1} = e^{-tA}$. From these facts we conclude that $e^{tA} \in H$ implies $e^{tA'} = e^{-tA}$. Taking derivatives on both sides at $t = 0$, we see that $A' = -A$. Conversely, if $A' = -A$, then $e^{tA}(e^{tA})' = e^{tA}e^{tA'} = e^{tA}e^{-tA} = I$, which means that e^{tA} is an orthogonal matrix. Let $\mathfrak{o}(n)$ denote the linear subspace of skew symmetric matrices in $\mathscr{M}_n(\mathbf{R})$; it has dimension $n(n-1)/2$. Since $X \to e^X$ has nonsingular Jacobian at $0 \in \mathscr{M}_n(\mathbf{R})$, it is a diffeomorphism near O. The same will then be true of its restriction to $\mathfrak{o}(n)$: A neighborhood of $O \in \mathfrak{o}(n)$ maps diffeomorphically onto a neighborhood of $I \in O(n)$. This proves the following:

The homomorphism $t \to e^{tA}$ is a one-parameter subgroup of $O(n)$ if and only if $A' = -A$, which is the necessary and sufficient condition on $A = (a_{ij})$ in order that the tangent vector $\sum_{i,j} a_{ij}(\partial/\partial x_{ij})_e$ to $Gl(n, \mathbf{R})$ at the identity be tangent to the subgroup $O(n)$. The dimension of $O(n)$ is $n(n-1)/2$.

Finally we recall that if G is a Lie group and $Z \in T_e(G)$, then Z determines uniquely a one-parameter subgroup which we denoted earlier by $F(t, Z)$. We use this to define an exponential mapping on an arbitrary Lie group.

(6.8) Definition *The exponential mapping*, $\exp\colon T_e(G) \to G$, is defined by the formula $\exp Z = F(1, Z)$.

According to Theorem 5.13 we have the following properties:

(6.9) Theorem *For any Lie group G the mapping* $\exp\colon T_e(G) \to G$ *is C^∞ and $F(t) = \exp tZ$ is the unique one-parameter subgroup such that $\dot{F}(0) = Z$. The Jacobian matrix at* 0 *of* exp *is the identity, that is, at e,* $\exp_*$ is the identity.† *Finally, if G is a subgroup of $Gl(n, \mathbf{R})$, then for each $Z \in T_e(G)$ there is an $A = (a_{ij}) \in \mathscr{M}_n(\mathbf{R})$ such that $Z = \sum a_{ij}(\partial/\partial x_{ij})_e$, and for this Z,* $\exp tZ = e^{tA}$.

Exercises

1. Complete the proof that when A, B are commuting $n \times n$ matrices, then $e^{A+B} = e^A e^B$. [*Hint:* first prove this for the exponential function on $\mathbf{R}$, using multiplication of power series; then try a similar proof.]
2. Check directly that the mapping $t \to e^{tA}$ of Example 6.4 is a group homomorphism.

† This requires that we identify the tangent space to $T_e(G)$ at $Z = 0$ with $T_e(G)$ itself, a common practice when working with vector spaces; $\exp_*$ is the mapping of tangent spaces induced by exp.

3. Let $A = (a_{ij})$ be an $n \times n$ matrix such that $a_{ij} = 0$ if $j \leq i$. Prove that the one-parameter subgroup e^{tA} is not a circle group in $Gl(n, \boldsymbol{R})$.
4. Find the one-parameter subgroups of $Gl(2, \boldsymbol{R})$ corresponding to A and B with

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Find the corresponding actions on $\boldsymbol{R}^2$ and their infinitesimal generators, starting from the natural action of $Gl(2, \boldsymbol{R})$ on $\boldsymbol{R}^2$.
5. Show that for any Lie group G the mapping $\mu: G \to G$ defined by $\mu(g) = g^{-1}$ takes left invariant vector fields to right invariant vector fields.
6. Prove that if A is a nonsingular $n \times n$ matrix and $X \in \mathscr{M}_n(\boldsymbol{R})$, then $Ae^X A^{-1} = e^{AXA^{-1}}$. From this deduce that $\det e^X = e^{\operatorname{tr} X}$. Use this to determine those matrices A such that e^{tA} is a one-parameter subgroup of $Sl(n, \boldsymbol{R})$.
7. Using the coordinate frames $\partial/\partial x_{ij}$, $i \leq i, j \leq n$ on $Gl(n, \boldsymbol{R})$, show that the vector field Z on $Gl(n, \boldsymbol{R})$ whose matrix of components at the identity is $A = (a_{ij})$ and XA at the element $X = (x_{ij})$ of $Gl(n, \boldsymbol{R})$ is a left-invariant vector field.

7 The Lie Algebra of Vector Fields on a Manifold

We denote by $\mathfrak{X}(M)$ the set of all C^∞-vector fields defined on the C^∞ manifold M. It is itself a vector space over $\boldsymbol{R}$ since if X and Y are C^∞-vector fields on M so is any linear combination of them with constant coefficients. In fact any linear combination with coefficients which are C^∞ functions on M is again a C^∞-vector field. For $X, Y \in \mathfrak{X}(M)$ and $f, g \in C^\infty(M)$ implies that the vector field $Z = fX + gY$, with the obvious definition $Z_p = f(p)X_p + g(p)Y_p$ for each $p \in M$ is a C^∞-vector field. We may express this as follows:

$\mathfrak{X}(M)$ is a *vector space* over $\boldsymbol{R}$ and a *module* over $C^\infty(M)$.

As a vector space $\mathfrak{X}(M)$ is not finite-dimensional over $\boldsymbol{R}$ (Exercise 1). In fact $\mathfrak{X}(M)$ is something more than just a vector space as we shall see.

(7.1) Definition We shall say that a vector space $\mathscr{L}$ over $\boldsymbol{R}$ is a (real) *Lie algebra* if in addition to its vector space structure it possesses a product, that is, a map $\mathscr{L} \times \mathscr{L} \to \mathscr{L}$, taking the pair (X, Y) to the element $[X, Y]$ of $\mathscr{L}$, which has the following properties:

(1) it is bilinear over $\boldsymbol{R}$:

$$[\alpha_1 X_2 + \alpha_2 X_2, Y] = \alpha_1[X_1, Y] + \alpha_2[X_2, Y],$$
$$[X, \alpha_1 Y_1 + \alpha_2 Y_2] = \alpha_1[X, Y_1] + \alpha_2[X, Y_2];$$

(2) it is skew commutative:

$$[X, Y] = -[X, Y];$$

(3) it satisfies the *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

(7.2) Example A vector space V^3, of dimension 3 over $\boldsymbol{R}$ with the usual vector product of vector calculus is a Lie algebra.

(7.3) Example Let $\mathscr{M}_n(\boldsymbol{R})$ denote the algebra of $n \times n$ matrices over $\boldsymbol{R}$ with XY denoting the usual matrix product of X and Y. Then $[X, Y] = XY - YX$, the "commutator" of X and Y, defines a Lie algebra structure on $\mathscr{M}_n(\boldsymbol{R})$ as is easily verified.

Now suppose that X and Y denote C^∞-vector fields on a manifold M, that is, $X, Y \in \mathfrak{X}(M)$. Then, in general, the operator $f \to X_p(Yf)$ defined on $C^\infty(p)$—f being a C^∞ function on a neighborhood of p—does *not* define a vector at p. Thus XY, considered as an operator on C^∞ functions on M, does not in general determine a C^∞-vector field. However, oddly enough, $XY - YX$ does; it defines a vector field $Z \in \mathfrak{X}(M)$ according to the prescription

$$Z_p f = (XY - YX)_p f = X_p(Yf) - Y_p(Xf), \qquad f \in C^\infty(p).$$

For if $f \in C^\infty(p)$, then Xf and Yf are C^∞ on a neighborhood of p, and this prescription determines a linear map of $C^\infty(p) \to \boldsymbol{R}$. Therefore if the Leibniz rule holds for Z_p, then Z_p is an element of $T_p(M)$ at each $p \in M$. Consider $f, g \in C^\infty(p)$. Then $f, g \in C^\infty(U)$ for some open set U containing p. Using the notation $(Xf)_p$ for $X_p f$, the value of Xf at p, we have the relations:

$$\begin{aligned}
(XY - YX)_p(fg) &= X_p(Yfg) - Y_p(Xfg) \\
&= X_p(fYg + gYf) - Y_p(fXg + gXf) \\
&= (X_p f)(Yg)_p + f(p)X_p(Yg) + (X_p g)(Yf)_p \\
&\quad + g(p)X_p(Yf) - (Y_p f)(Xg)_p - f(p)Y_p(Xg) \\
&\quad - (Y_p g)(Xf)_p - g(p)(Y_p Xf),
\end{aligned}$$

so that

$$\begin{aligned}
Z_p(fg) &= (XY - YX)_p(fg) = f(p)(XY - YX)_p g + g(p)(XY - YX)_p f \\
&= f(p)Z_p g + g(p)Z_p f.
\end{aligned}$$

Finally, if f is C^∞ on any open set $U \subset M$, then so is $(XY - YX)f$, and therefore Z is a C^∞-vector field on M as claimed.

We may define a product on $\mathfrak{X}(M)$ using this fact; namely, define the product of X and Y by $[X, Y] = XY - YX$.

(7.4) Theorem *$\mathfrak{X}(M)$ with the product $[X, Y]$ is a Lie algebra.*

Proof If $\alpha, \beta \in \mathbf{R}$ and X_1, X_2, Y are C^∞-vector fields, then it is straightforward to verify that

$$[\alpha X_1 + \beta X_2, Y]f = \alpha[X_1, Y]f + \beta[X_2, Y]f.$$

Thus $[X, Y]$ is linear in the first variable. Since the skew commutativity $[X, Y] = -[Y, X]$ is immediate from the definition, we see that linearity in the first variable implies linearity in the second. Therefore $[X, Y]$ is bilinear and skew-commutative. There remains the Jacobi identity which follows immediately if we evaluate $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]]$ applied to a C^∞-function f. Using the definition, we obtain

$$\begin{aligned}[X, [Y, Z]]f &= X(([Y, Z])f) - [Y, Z](Xf) \\ &= X(Y(Zf)) - X(Z(Yf)) - Y(Z(Xf)) + Z(Y(Xf)).\end{aligned}$$

Permuting cyclically and adding establishes the identity.

(7.5) Remark $[X, Y]$ is not $C^\infty(M)$ bilinear. In fact for $f \in C^\infty(M)$, $[X, fY] = f[X, Y] + (Xf)Y$ as is shown in Exercise 2. This may be used to derive a formula for the components of $[X, Y]$ in local coordinates (Exercise 3).

We now make use of a vector field X on M to define a method of differentiation which has many applications in manifold theory. We have already defined the derivative of a function $f \in C^\infty(M)$ at a point p in the direction of X; it is just $X_p f$. This generalizes from $\mathbf{R}^n$ to an arbitrary manifold the notion of directional derivative of a function. However, if we wish to determine the rate of change of a vector field Y at $p \in M$ in some direction X_p, we have trouble as soon as we leave $\mathbf{R}^n$, for it is only in $\mathbf{R}^n$ that we are able to compare the value of Y at p with its value at nearby points, which we must do to compute a rate of change. Now, given a vector field X on M, there is an associated one-parameter group $\theta\colon W \to M$ generated by X. For each $t \in R$ we know (Theorem 3.12) that $\theta_t\colon V_t \to V_{-t}$ is a diffeomorphism (with inverse θ_{-t}) of the open set V_t, provided V_t is not empty. In particular for each $p \in M$ there is a neighborhood V and a $\delta > 0$ such that $V \subset V_t$ for $|t| < \delta$. The isomorphism $\theta_{t*}\colon T_p(M) \to T_{\theta_t(p)}(M)$ and its inverse allow us to compare the values of vector fields at these two points.

Indeed, suppose Y is a second C^∞-vector field on M. We may use this idea to compute for each p the rate of change of Y in the direction of X, that is, along the integral curve of the vector field X passing through p. We shall denote this rate of change by $L_X Y$; it is itself a C^∞-vector field.

(7.6) Definition The vector field $L_X Y$, called the *Lie derivative* of Y with respect to X is defined at each $p \in M$ by either of the following limits.

$$(L_X Y)_p = \lim_{t \to 0} \frac{1}{t} [\theta_{-t*}(Y_{\theta(t,p)}) - Y_p]$$

$$= \lim_{t \to 0} \frac{1}{t} [Y_p - \theta_{t*} Y_{\theta(-t,p)}].$$

The second definition is obtained from the first by replacing t by $-t$. We interpret the first expression as follows: Apply to $Y_{\theta(t,p)} \in T_{\theta(t,p)}(M)$ the isomorphism θ_{-t*}, taking $T_{\theta(t,p)}(M)$ to $T_p(M)$. Then in $T_p(M)$ take the difference of this vector and Y_p, multiply by the scalar $1/t$, and pass to the limit as $t \to 0$. Equivalently, we may identify $Z_p(t) = \theta_{-t*}(Y_{\theta(t,p)}) \in T_p(M)$ with a curve of $\boldsymbol{R}^n$; then $(L_X Y)_p = \dot{Z}_p(0)$. The fact that the vector *field* so defined is C^∞ may be verified by writing the formula above in local coordinates (Exercise 6). We shall give another characterization of $L_X Y$ which requires a modification of Lemma II.4.3, following Kobayashi and Nomizu [1, p. 15].

(7.7) Lemma *Let X be a C^∞-vector field on M and θ be the corresponding map of $W \subset R \times M$ onto M. Given $p \in M$ and $f \in C^\infty(U)$, U an open set containing p, we choose $\delta > 0$ and a neighborhood V of p in U such that $\theta(I_\delta \times V) \subset U$. Then there is a C^∞ function $g(t, q)$ defined on $I_\delta \times V$ such that for $q \in V$ and $t \in I_\delta$ we have*

$$f(\theta_t(q)) = f(q) + tg(t, q) \qquad \textit{and} \qquad X_q f = g(0, q).$$

Proof There is a neighborhood V of p and a $\delta > 0$ such that $\theta_t(p) = \theta(t, p)$ is defined and C^∞ on $I_\delta \times V$ and maps $I_\delta \times V$ into U according to Theorem 4.2. The function $r(t, q) = f(\theta_t(q)) - f(q)$ is C^∞ on $I_\delta \times V$ and $r(0, q) = 0$. We denote by $\dot{r}(t, q)$ its derivative with respect to t. We define $g(t, q)$—for each fixed q—by the formula

$$g(t, q) = \int_0^1 \dot{r}(ts, q)\, ds.$$

This function is also C^∞ on $I_\delta \times V$ (verified by use of local coordinates and properties of the integral). By the fundamental theorem of calculus,

$$tg(t, q) = \int_0^1 \dot{r}(ts, q)t\, ds = r(t, q) - r(0, q) = r(t, q).$$

Using the definition of r, this becomes

$$f(\theta_t(q)) = f(q) + tg(t, q).$$

On the other hand, by the definition (3.2) of the infinitesimal generator of θ,

$$g(0, q) = \lim_{t \to 0} g(t, q) = \lim_{t \to 0} \frac{1}{t} r(t, q) = \lim_{t \to 0} \frac{1}{t} [f(\theta_t(q)) - f(q)] = X_q f. \quad \blacksquare$$

We use the lemma to prove the following theorem:

(7.8) Theorem *If X and Y are C^∞-vector fields on M, then*

$$L_X Y = [X, Y].$$

Proof For any $p \in M$ and any function f which is C^∞ near p, we have for some $\delta > 0$

$$\frac{1}{t} [Y_p - \theta_{t*}(Y_{\theta(-t, p)})] f = \frac{1}{t} [Y_p f - Y_{\theta(-t, p)}(f \circ \theta_t)]$$

for $0 < |t| < \delta$. If we can show that as $t \to 0$ the limit of the right-hand side is $[X, Y]_p f$ for all such f, then $(L_X Y)_p$ is defined: $(L_X Y)_p f = [X, Y]_p f$ for all f; and thus $(L_X Y)_p = [X, Y]_p$. So the conclusion of the theorem follows. Making use of Lemma 7.7 and setting $g_t(q) = g(t, q)$, we are led to consider then

$$\lim_{t \to 0} \frac{1}{t} [Y_p f - Y_{\theta(-t, p)}(f + t g_t)].$$

Collecting the first two terms and replacing t by $-t$ yields

$$\lim_{t \to 0} \frac{1}{t} [(Yf)(\theta_t(p)) - (Yf)(p)] - \lim_{t \to 0} Y_{\theta_t(p)} g_t \,.$$

By formula (3.2) with f replaced by Yf and Δt by t, the first term is $X_p(Yf)$. Since $\theta_t(p) \to p$ and $g_t(q) \to X_q f$ as $t \to 0$, $Y_{\theta_t(p)} g_t \to Y_p(Xf)$ and we obtain in the limit

$$X_p(Yf) - Y_p(Xf) = [X, Y]_p f.$$

This completes the proof of the theorem; it also shows that $L_X Y$ is C^∞. $\blacksquare$

(7.9) Theorem *Let $F: N \to M$ be a C^∞ mapping and suppose that X_1, X_2 and Y_1, Y_2 are vector fields on N, M, respectively, which are F-related, that is, for $i = 1, 2$, $F_*(X_i) = Y_i$. Then $[X_1, X_2]$ and $[Y_1, Y_2]$ are F-related, that is, $F_*[X_1, X_2] = [F_*(X_1), F_*(X_2)]$.*

Proof Before proving the theorem we note the following necessary and sufficient condition for X on N and Y on M to be F-related: for any g which is C^∞ on some open set $V \subset M$,

$$(*) \qquad (Yg) \circ F = X(g \circ F)$$

on $F^{-1}(V)$. This is essentially a restatement of the definition of F-related, for if $q \in F^{-1}(V)$, then $F_*(X_q)g = X_q(g \circ F) = X(g \circ F)(q)$; and $Y_{F(q)}g$ is the value of the C^∞ function Yg at $F(q)$, that is, $((Yg) \circ F)(q)$. Thus (*) holds for all g if and only if $F_*(X_q) = Y_{F(q)}$ for all $q \in M$.

Returning to the proof we consider $f \in C^\infty(V)$, $V \subset M$, so that $Y_1 f$ and $Y_2 f \in C^\infty(V)$ also. Apply (*), first with $g = Y_2 f$ and then with $g = f$ giving the equalities

$$[Y_1(Y_2 f)] \circ F = X_1((Y_2 f) \circ F) = X_1[X_2(f \circ F)].$$

Interchanging the roles of Y_1, Y_2 and X_1, X_2 and subtracting, we obtain

$$([Y_1, Y_2]f) \circ F = [X_1, X_2](f \circ F),$$

which by (*) is equivalent to $[X_1, X_2]$ and $[Y_1, Y_2]$ being F-related. ∎

We now define the Lie algebra $\mathfrak{g}$ of a Lie group G.

(7.10) Corollary *If G is a Lie group, then the left-invariant vector fields on G form a Lie algebra $\mathfrak{g}$ with the product $[X, Y]$ and $\dim \mathfrak{g} = \dim G$. If $F: G_1 \to G_2$ is a homomorphism of Lie groups, $F_*: \mathfrak{g}_1 \to \mathfrak{g}_2$ is a homomorphism of Lie algebras.*

Proof Let $a \in G$, and let X and Y be left-invariant vector fields. L_a (left translation) is a diffeomorphism and $L_{a*}X = X$, $L_{a*}Y = Y$. Therefore $L_{a*}[X, Y] = [X, Y]$ by the theorem, so $[X, Y]$ is L_a-invariant for any a. Hence the subspace $\mathfrak{g}$ of left-invariant vector fields is closed with respect to $[X, Y]$. Since each $X \in \mathfrak{g}$ is uniquely determined by X_e, the mapping $X \to X_e$ is an isomorphism of $\mathfrak{g}$ and $T_e(G)$ as vector spaces. The last statement follows from Corollary 2.10 and Theorem 7.9. ∎

(7.11) Remark If $H \subset G$ is a Lie subgroup, then Corollary 7.10 implies that $i_*(\mathfrak{h})$ is a subalgebra of $\mathfrak{g}$. It consists of the elements of $\mathfrak{g}$ tangent to H and its cosets gH.

(7.12) Theorem *Let X and Y be C^∞ vector fields on a manifold M and let θ, σ denote the local one-parameter group actions they generate. Then $[X, Y] \equiv 0$ on M if and only if for each $p \in M$ there is a $\delta_p > 0$ such that $\sigma_s \circ \theta_t(p) = \theta_t \circ \sigma_s(p)$ for $|t|, |s| < \delta_p$.*

Proof If $p \in M$, then there exists a neighborhood V of p and $\delta > 0$ such that $\sigma_s \circ \theta_t(q)$ and $\theta_t \circ \sigma_s(q)$ are defined for $q \in V$ and $|s|, |t| < \delta$. If $\sigma_s \circ \theta_t(q) =$

$\theta_t \circ \sigma_s(q)$ on this set, then by Theorem (5.7), Y is θ_t-invariant on V for any fixed small t, i.e., $\theta_{t*}(Y_q) = Y_{\theta(t,q)}$. It follows that

$$[X, Y]_q = (L_X Y)_q = \lim_{t\to 0} \frac{1}{t} [\theta_{-t*}(Y_{\theta(t,q)}) - Y] = \lim_{t\to 0} \frac{1}{t} [Y_q - Y_q] = 0.$$

Conversely, suppose $[X, Y] \equiv 0$ on M. Given $q \in V$, we let $Z_q(t) = \theta_{-t*}(Y_{\theta(t,q)})$ and show $\dot{Z}_q(t) = 0$. Let $q' = \theta(t, q)$; then

$$\begin{aligned}\dot{Z}_q(t) &= \lim_{\Delta t\to 0} \frac{1}{\Delta t} [\theta_{-(t+\Delta t, q)*}(Y_{\theta(\Delta t, q')}) - \theta_{-t*}(Y_{q'})] \\ &= \theta_{-t*}\left(\lim \frac{1}{\Delta t}[\theta_{-\Delta t*}(Y_{\theta(\Delta t, q')} - Y_{q'})]\right) \\ &= \theta_{-t*}((L_X Y)_{q'}) = 0,\end{aligned}$$

since $L_X Y = [X, Y] = 0$. Hence $Z_q(t)$ is constant for $|t| < \delta$ and $q \in V$ which means that Y is θ_t-invariant, and hence, by Theorem 5.7, $\theta_t \circ \sigma_s = \sigma_s \circ \theta_t$ on V.

Exercises

1. Show that $\mathfrak{X}(M)$ is infinite-dimensional over R but locally finitely generated over $C^\infty(M)$, that is, each $p \in M$ has a neighborhood V on which there is a finite set of vector fields which generate $\mathfrak{X}(M)$ as a $C^\infty(V)$ module.
2. Let $X, Y \in \mathfrak{X}(M)$ and $f, g \in C^\infty(M)$ and prove that

$$[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X.$$

3. Suppose U, φ is a coordinate neighborhood on M, $X, Y \in \mathfrak{X}(M)$, and $E_1, \ldots, E_n$ the coordinate frames, and note that $[E_i, E_j] = 0$ on U. If $X = \sum_i \alpha^i E_i$ and $Y = \sum_j \beta^j E_j$ on U, then show that

$$[X, Y] = \sum_{i,j} \left(\alpha^i \frac{\partial \beta^j}{\partial x^i} - \beta^i \frac{\partial \alpha^j}{\partial x^i}\right) E_j \qquad \text{on} \quad U.$$

4. Given the vector fields in $\boldsymbol{R}^3$ (with coordinates x, y, z)

$$X = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}, \qquad Y = z\frac{\partial}{\partial y} - y\frac{\partial}{\partial z}, \qquad Z = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z},$$

compute the components of the three pairs of products.
5. Show that $(L_X Y)_p$ depends on the fact that we use vector *fields*, that is, if X, X' agree at p but are not the same as vector fields, then $(L_X Y)_p$ may differ from $(L_{X'} Y)_p$.

6. Write the expressions defining $(L_X Y)_p$ (Definition 7.6) in local coordinates of a coordinate neighborhood U, φ and show that $L_X Y$ is a C^∞-vector field on U.
7. Let $G = Gl(n, \mathbf{R})$ with matrix entries as local coordinates. To each $X \in \mathfrak{g}$ we assign the $n \times n$ matrix $A = (a_{ij})$ of components of X_e, $X_e = \sum_{i,j} a_{ij}(\partial/\partial x_{ij})_e$. Denote this mapping of $\mathfrak{g}$ onto $\mathcal{M}_n(R)$ by μ. Show that μ is an isomorphism to the algebra of Example 7.3, i.e.,

$$\mu[X, Y] = \mu(X)\mu(Y) - \mu(Y)\mu(X).$$

8. Show that whenever H is a (Lie) subgroup of the Lie group G, then the Lie algebra of H may be naturally identified with a subalgebra of $\mathfrak{g}$, thus verifying Remark 7.11.
9. Show that two one-parameter subgroups $F(t)$ and $G(t)$ commute elementwise if and only if their Lie algebras, in the sense of Exercise 8, satisfy $[X, Y] = 0$ for each X, Y in the algebra of $F(t)$, $G(t)$, respectively.
10. If $F\colon M \to N$ is a diffeomorphism of M onto N and X, Y are C^∞-vector fields on M, then prove that $F_*(L_X Y) = L_{F_*(X)} F_*(Y)$, that is, $L_X Y$ is F-related to $L_{F_*(X)} F_*(Y)$.
11. If $f \in C^\infty(M)$, $X, Y \in \mathfrak{X}(M)$, let $L_X f = Xf$ (cf. (3.2) and (7.6)). Show that $L_X(fY) = (L_X f)Y + f(L_X Y)$.

8 Frobenius' Theorem

The concept of vector fields on a manifold can be used to give a coordinate-free treatment of certain first-order linear partial differential equations which is useful even for local questions in $\mathbf{R}^n$ and indispensable in many global questions. First consider an example.

(8.1) Example We consider the system of partial differential equations

$$(*) \qquad \frac{\partial z}{\partial x} = g(x, y, z), \qquad \frac{\partial z}{\partial y} = h(x, y, z)$$

defined on some open subset W of $\mathbf{R}^3$. Given $(a, b, c) \in W$, a solution, if any exists, determined by (a, b, c) will be a function $z = f(x, y)$ such that $c = f(a, b)$ and $f_x(x, y) \equiv g(x, y, f(x, y))$, $f_y(x, y) \equiv h(x, y, f(x, y))$. Geometrically, a solution can be thought of as a surface through (a, b, c) whose tangent plane at any of its points has the vectors

$$X = \frac{\partial}{\partial x} + g(x, y, z)\frac{\partial}{\partial z}, \qquad Y = \frac{\partial}{\partial y} + h(x, y, z)\frac{\partial}{\partial z}$$

as a basis—these being orthogonal to the normal vector

$$N = f_x \frac{\partial}{\partial x} + f_y \frac{\partial}{\partial y} - \frac{\partial}{\partial z}$$

to the surface $z = f(x, y)$ above. It becomes clear however that for *arbitrary* independent vector fields X and Y, even of the form above, there may fail to be a solution. If there is a solution through any point of W, then X, Y can be written as

$$X = \frac{\partial}{\partial x} + f_x(x, y) \frac{\partial}{\partial z}, \qquad Y = \frac{\partial}{\partial y} + f_y(x, y) \frac{\partial}{\partial z}$$

and

$$[X, Y] = \left(\frac{\partial^2 f}{\partial x \, \partial y} - \frac{\partial^2 f}{\partial x \, \partial y} \right) \frac{\partial}{\partial z} = 0.$$

This is, then, a necessary condition for an everywhere independent pair of vector fields on W to determine a surface whose tangent plane they define. It reflects independence of order of differentiation in second derivatives.

(8.2) Some Definitions Let M be a manifold of dimension $m = n + k$ and assume that to each $p \in M$ is assigned an n-dimensional subspace Δ_p of $T_p(M)$. Suppose moreover that in a neighborhood U of each $p \in M$ there are n linearly independent C^∞-vector fields $X_1, \ldots, X_n$ which form a basis of Δ_q for every $q \in U$. Then we shall say that Δ is a C^∞ n-plane *distribution* of dimension n on M and $X_1, \ldots, X_n$ is a *local basis* of Δ.

We shall say that the distribution Δ is *involutive* if there exists a local basis $X_1, \ldots, X_n$ in a neighborhood of each point such that

$$[X_i, X_j] = \sum_{k=1}^{n} c_{ij}^k X_k, \qquad 1 \le i, \quad j \le n.$$

[The c_{ij}^k will not in general be constants, but will be C^∞ functions on the neighborhood.] Exercises 1 and 2 clarify this concept.

Finally, if Δ is a C^∞ distribution on M and N is a connected C^∞ submanifold of M such that for each $q \in N$ we have $Tq(N) \subset \Delta g$, then we shall say that N is an *integral manifold* of Δ. Note that an integral manifold may be of lower dimension than Δ, and need not be a regular submanifold.

An example of an involutive distribution is the following: Let $M = \boldsymbol{R}^{n+k}$ and let Δ be spanned by $X_i = \partial/\partial x^i$, $i = 1, \ldots, n$. Then the distribution is the subspace of dimension n consisting of all those vectors parallel to $\boldsymbol{R}^n$ at each point q of M. We shall see that this apparently rather special example is actually typical, locally, of involutive distributions.

Let Δ be a C^∞ distribution on M of dimension n, the dimension of M being $m = n + k$. We shall say that Δ is *completely integrable* if each point $p \in M$ has a cubical coordinate neighborhood U, φ such that if $x^1, \ldots, x^m$ denote the local coordinates, then the n vectors $E_i = \varphi_*^{-1}(\partial/\partial x^i)$, $i = 1, \ldots, n$, are a local basis on U for Δ. Note that in this case there is an n-dimensional integral manifold N through each point q of U such that $T_q(N) = \Delta_q$, i.e., dim $N = n$. In fact, if $(a^1, \ldots, a^m)$ denote the coordinates of q, then an integral manifold through q is the n-slice defined by $x^{n+1} = a^{n+1}, \ldots, x^m = a^m$, that is, $N = \varphi^{-1}\{x \in \varphi(U) | x^j = a^j,\ j = n+1, \ldots, m\}$, a *slice* of U. Of course, in this case the distribution is involutive for

$$[E_i, E_j] = \varphi_*^{-1}\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0, \qquad 1 \le i, \quad j \le n.$$

We shall call U, φ *flat* with respect to Δ. Thus complete integrability is equivalent to every point having a flat coordinate neighborhood. Thus any completely integrable distribution is involutive (for a generalization, see Exercise 4). However, most distributions are not involutive, for example, on $\boldsymbol{R}^3$ the distribution

$$X_1 = x^3 \frac{\partial}{\partial x^1} + \frac{\partial}{\partial x^3}, \qquad X_2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial x^3}$$

is not involutive since $[X_1, X_2] = -\partial/\partial x^1$, which is not a linear combination of X_1 and X_2.

An important and instructive example of an involutive distribution is furnished by the Lie algebra $\mathfrak{h}$ of a subgroup H of a Lie group G; $\mathfrak{h}$ consists of left-invariant vector fields on G which are tangent to H at the identity. As we have seen (Remark 7.11), this determines a subalgebra, the image of the Lie algebra of H under the inclusion mapping. These give a (left-invariant) distribution Δ on G such that $\Delta_h = T_h(H)$ for every $h \in H$. The cosets gH are the integral manifolds of this distribution—which is evidently involutive since $\mathfrak{h}$ is a subalgebra of $\mathfrak{g}$.

A distribution Δ of dimension 1 is just a field of line elements, that is, one-dimensional subspaces. A local basis is given by any nonvanishing vector field X which belongs to Δ at each point and, of course, an integral curve of X is an integral manifold. We know from the existence theorem that such integral manifolds passing through any given point exist and are unique. In fact, Theorem 3.14 says precisely that any such distribution is completely integrable. It is also involutive since $[X, X] = 0$. In the light of these remarks, the following theorem may be considered a generalization of the existence theorem (Theorem 4.1) to certain types of partial differential equations. In the general case, as we have seen, there is a necessary condition which is not automatic—as it is in the case of a one-dimensional distribution. This condition is the involutivity of Δ.

(8.3) Theorem (Frobenius) *A distribution Δ on a manifold M is completely integrable if and only if it is involutive.*

Proof We showed above that a completely integrable distribution is involutive. This is an easy consequence of the definitions. We shall prove that involutive distributions are completely integrable by first proving what appears to be a special case. Assume throughout that $m = \dim M$ and $n = \dim \Delta$.

Lemma *Let $X_1, \ldots, X_n$ be C^∞ vector fields defined on an open subset $U \subset \boldsymbol{R}^m$ such that* (i) *$\{X_{1q}, \ldots, X_{nq}\}$ are linearly independent at each $q \in U$ and* (ii) *$[X_i, X_j] \equiv 0$ on U, $1 \leq i, j \leq n$. Then there is a coordinate neighborhood V, ψ around any $p \in U$ such that $X_i = E_i = \psi_*^{-1}(\partial/\partial x^i)$, $i = 1, \ldots, n$, on V.*

Proof Given $p \in U$, we may assume coordinates $(x^1, \ldots, x^m)$ in $\boldsymbol{R}^m$ such that $p = (0, .., 0)$ and $X_{ip} = (\partial/\partial x^i)_0$ for $i = 1, \ldots, n$. Let $\theta_t^i(q) = \theta^i(t, q)$ denote the one-parameter group action generated by X_i. Using repeatedly both the existence theorem and continuity, we may assume that there exists a neighborhood $W \subset U$ of p and a $\delta > 0$ such that for $|t_i| < \delta$ and for any permutation σ of $\{1, \ldots, n\}$ the mapping $\theta_{t_{\sigma(1)}}^1 \circ \cdots \circ \theta_{t_{\sigma(n)}}^n$ is defined on W. Since all $[X_i, X_j] = 0$ on U, Theorem 7.12 implies that $\theta_{t_{\sigma(1)}}^{\sigma(1)} \circ \cdots \circ \theta_{t_{\sigma(n)}}^{\sigma(n)}(q) = \theta_{t_1}^1 \circ \cdots \theta_{t_n}^n(q)$ for any $q \in W$ and $\sigma \in \mathfrak{S}_n$, the symmetric group of degree n, i.e., all permutations of $\{1, \ldots, n\}$. Let $W' \subset W$ be a neighborhood of the origin such that $x = (x^1, \ldots, x^n, x^{n+1}, \ldots, x^m) \in W'$ implies that $|x^i| < \delta$, $i = 1, \ldots, n$. We define a C^∞ mapping $\theta \colon W' \to \boldsymbol{R}^m$ by

$$\theta(x^1, \ldots, x^n, \ldots, x^m) = \theta_{x^1}^1 \circ \cdots \circ \theta_{x^n}^m(0, \ldots, 0, x^{n+1}, \ldots, x^m).$$

In other words, we project x to the plane of the last $m - n$ coordinates and then act by the various one-parameter groups, i.e., move along the various integral curves of the X_i. This may be done in *any order* with the same results. Note that $\theta(0, \ldots, 0, x^{n+1}, \ldots, x^m) = (0, \ldots, 0, x^{n+1}, \ldots, x^m)$, i.e., θ is the identity on the $(m - n)$-slice $x^i = 0$, $i = 1, \ldots, n$, of W'.

We now consider $\theta_*(\partial/\partial x^i|_a)$, $a \in W'$. For given i and a, let $F(t) = \theta(a^1, \ldots, a^{i-1}, t, a^{i+1}, \ldots, a^m)$; then $\theta_*(\partial/\partial x^i|_a) = \dot{F}(a^i)$. First note that for $i > n$ and $a = (0, \ldots, 0, a^{n+1}, \ldots, a^m)$, $\theta_*(\partial/\partial x^i) = \partial/\partial x^i$. In particular, this holds at $p = (0, \ldots, 0)$. For $i \leq n$ and any $a \in W'$,

$$F(t) = \theta_t^i[\theta_{a^1}^1 \circ \cdots \circ \hat{\theta}_t^i \circ \cdots \circ \theta_{a^n}^n(0, \ldots, 0, a^{n+1}, \ldots, a^m)],$$

where the caret denotes that θ_t^i is omitted. We have used the independence of order. Let q denote the point in brackets; then $F(t) = \theta_t^i(q)$, so $\dot{F}(a^i) = X_{i\theta^i(a^i, q)} = X_{i\theta(a, q)}$. Thus $\theta_*(\partial/\partial x^i|_a) = X_{i\theta(a)}$, and in particular then $\theta_*(\partial/\partial x^i|_0) = X_{ip} = \partial/\partial x^i|_0$ for $i = 1, \ldots, n$. It follows that θ is a diffeomorphism on a neighborhood $V' \subset W'$ of the origin and that $V = \theta(V')$, $\psi = \theta^{-1}$ is the desired coordinate neighborhood. ∎

We now proceed to the proof of the theorem in two steps. First we show that in a neighborhood U of any $p \in M$ an involutive n-plane distribution Δ has a local basis of vector fields $Y_1, \ldots, Y_n$ satisfying $[Y_i, Y_j] \equiv 0$. We begin with an arbitrary local basis $X_1, \ldots, X_n$ defined, we may suppose, on a coordinate neighborhood U', φ of p. Relative to the coordinate frames $X_{iq} = \sum_{a=1}^m \alpha_{ia}(q) E_{aq}$ and we may suppose coordinates so numbered that the matrix $(\alpha_{ij}(q))$, $1 \le i, j \le n$ is nonsingular on a neighborhood $U \subset U'$ of p. Let $(\beta_{ij}(q))$ denote its inverse and define a new basis of Δ on U by $Y_{iq} = \sum_{j=1}^n \beta_{ij}(q) X_{iq} = E_{iq} + \sum_{\nu=n+1}^m \gamma_{i\nu}(q) E_{\nu q}$, where $\gamma_{i\nu} = \sum_{j=1}^n \beta_{ij}\alpha_{j\nu}$, $i = 1, \ldots, n$; $\nu = n+1, \ldots, m$. Since Δ is involutive $[Y_i, Y_j] = \sum_{k=1}^n c_{ij}^k \; Y_k = \sum_{k=1}^n c_{ij}^k(E_k + \sum_{\tau=n+1}^m \gamma_{k\tau} E_t)$ on U. On the other hand, $[Y_i, Y_j] = [E_i + \sum_\nu \gamma_{i\nu} E_\nu, E_j + \sum_\mu \gamma_{j\mu} E_\mu] = \sum_{\tau=n+1}^m \sigma_{ij}^\tau E_\tau$. Since $[E_a, E_b] = 0$ for $1 \le a, b \le m$. It is possible for these two linear combinations of $E_1, \ldots, E_m$ to be equal only if $c_{ij}^k = 0$ on U for $1 \le i, j, k \le n$, so $Y_1, \ldots, Y_n$ is the desired local basis.

The second step of the proof is to simply put things together. Assuming Δ is an involutive n-plane distribution on M and $p \in M$, we choose a coordinate neighborhood W, σ of p on which there is a local basis $Y_1, \ldots, Y_n$ such that $[Y_i, Y_j] \equiv 0$. In $\sigma(W) \subset R^m$ the vector fields $\tilde{Y}_i = \sigma_*(Y_i)$ have the same property so that by the lemma there exists a coordinate neighborhood V, ψ of $\sigma(p)$, $V \subset \sigma(W)$ with coordinates $y^1, \ldots, y^m$ and with $\tilde{Y}_i = \psi_*^{-1}(\partial/\partial y^i)$, for $i = 1, \ldots, n$. It is no loss of generality to assume $\psi(V) = C_\varepsilon^m(0)$ and $\psi(\sigma(p)) = (0, \ldots, 0)$. We now define the coordinate neighborhood $\tilde{U}$, φ of p on M by $\tilde{U} = \sigma^{-1}(V)$ with $\varphi = \psi \circ \sigma$. Then $\varphi(p) = (0, \ldots, 0)$, $\varphi(\tilde{U}) = C_\varepsilon^m(0)$ and $E_i = \sigma_*^{-1} \circ \psi_*^{-1}(\partial/\partial y^i) = \sigma_*^{-1}(\tilde{Y}_i) = Y_i$ for $i = 1, \ldots, n$. This proves that Δ is completely integrable.

Theorem 8.3 implies the following corollary which is essentially a local uniqueness statement for integral manifolds of an involutive n-dimensional distribution Δ in a manifold M of dimension m.

(8.4) Corollary *Let U, φ be a flat coordinate neighborhood relative to the involutive n-plane distribution Δ on M. Then any connected integral manifold $C \subset U$ must lie on a single n-slice*

$$S_a = \{q \in U \,|\, x^{n+1}(q) = a^{n+1}, \ldots, x^m(q) = a^m\}.$$

Proof It is clearly sufficient to show that this holds for any integral curve $t \to F(t) \subset U$, $a < t < b$, since any two points of C lie on such a curve. By definition $\dot{F}(t) \in \Delta_{F(t)}$ so $\dot{F}(t) = \sum_{i=1}^n \alpha^i(t) E_{iF(t)}$. Consider $(d/dt)y^j(F(t))$ for $j > n$. We have

$$\frac{d}{dt}\, y^j(F(t)) = F_*\left(\frac{d}{dt}\right)y^j = \sum_{i=1}^n \alpha^i(t) E_{iF(t)} y^j = 0$$

since $E_i y^j = 0$ unless $i = j$. Since the curve is connected, $y^j(F(t)) = a^j$, a constant for $j = n+1, \ldots, m$, so $F(t)$ lies on S_a. ∎

The following result should be compared with Lemma III.6.7.

(8.5) Theorem *Let $N \subset M$ be an integral manifold of an involutive distribution Δ with* $\dim N = \dim \Delta$ *and suppose $F: A \to M$ is a C^∞ mapping of a manifold A into M. If $F(A) \subset N$, then F is C^∞ as a mapping into N.*

Proof Let $F_1: A \to N$ denote the mapping of A into N defined by F. Then $F = i \circ F_1$, where $i: N \to M$ is the inclusion map—a one-to-one, C^∞ immersion. We are to prove that F_1 is C^∞. If $p \in A$, let $q = F_1(p) = F(p)$ and let U, φ be a flat coordinate neighborhood of q with respect to Δ. Since $i: N \to M$ is continuous, $N \cap U$ is open in N and thus the union of a countable collection of disjoint connected open sets, each—as connected integral manifolds—lying on an n-slice of U. Let V be that one which contains q. Since $\dim N = n$, V is an open subset of the n-slice through q and its image $\varphi(V)$ is an open subset of the slice $S_0 = \{y \in C_\varepsilon^m(0) \mid y^{n+1} = \cdots = y^m = 0\}$ of $\varphi(U)$. By choosing $\varepsilon' < \varepsilon$, we may assume that $\varphi(V) \cap C_{\varepsilon'}^m(0)$ is all of the n-slice through the origin. Let $U' = \varphi^{-1}(C_{\varepsilon'}^m(0)$ and $V' = V \cap U'$. Then U', $\varphi' = \varphi | U'$ is flat and V' is an open neighborhood of N which is exactly an n-slice of the flat neighborhood U'. Let π denote the projection, $\pi(y^1, \ldots, y^m) \to (y^1, \ldots, y^n)$. Then V', $\psi' = \pi \circ \varphi'$ is a coordinate neighborhood of q on N. The inclusion i is given in these local coordinates by $\hat{\imath}: (y^1, \ldots, y^n) \to (y^n, \ldots, y^n, 0, \ldots, 0)$.

Now let W, θ be a connected coordinate neighborhood of p on A such that $F(W) \subset U'$. Since $F(W) = i \circ F_1(W) \subset N \cap U'$, it must lie on a single slice; but $q \in F(W)$, so this is the slice V'. In local coordinates $(x^1, \ldots, x^r)$ on W, $\hat{F}(x^1, \ldots, x^r) = (f^1(x), \ldots, f^n(x), 0, \ldots, 0)$ with f^i being C^∞. Thus F_1 in local coordinates of W, θ and V', ψ' is given by $\hat{F}_1(x^1, \ldots, x^r) = (f^1(x), \ldots, f^n(x))$. It follows that F_1 is C^∞ on the neighborhood W of p. Since p was arbitrary this proves the theorem. ∎

(8.6) Definition A *maximal integral manifold* N of an involutive distribution Δ is a connected integral manifold which contains every connected integral manifold which has a point in common with it.

It is immediate from Corollary 8.4 that a maximal integral manifold has the same dimension as Δ. It is also clear that at most one maximal integral manifold can pass through a point p of M. It is true but more difficult to prove that there is a maximal integral manifold through every point of M. The idea is to piece together local slices using Corollary 8.4 and build up an immersed submanifold. The difficulty is in showing that there are not too many such slices, that is, in proving that we have a countable basis of open sets. We shall not prove this here. Proofs are given by Chevalley [1] and Warner [1], for example. We shall state one theorem whose proof requires this fact and show how it is used.

(8.7) Theorem *Let G be a Lie group, $\mathfrak{g}$ its Lie algebra, and $\mathfrak{h}$ a subalgebra of $\mathfrak{g}$ Then there is a unique connected subgroup H of G whose Lie algebra is $\mathfrak{h}$.*

Proof Let the left-invariant vector fields $X_1, \ldots, X_n$ on $\mathfrak{g}$ be a basis of $\mathfrak{h}$. They define an involutive distribution Δ which is invariant under left translations; hence each integral manifold N is carried by any left translation. L_g diffeomorphically to an integral manifold $L_g(N)$. Let H be the maximal integral manifold through the identity element e. If $h \in H$, then $L_{h^{-1}}(h) = e$ so that both H and $L_{h^{-1}}(H)$ have e in common. Since H is maximal, $L_{h^{-1}}(H) = H$. It follows that if $h_1, h_2 \in H$, then $h_1^{-1}h_2 \in H$ and H is thus a subgroup as well as an immersed submanifold. The product mapping $H \times H \to H$ is a composition of inclusion $i: H \times H \to G \times G$ and the product mapping $\theta: G \times G \to G$. Both are C^∞ so that $\theta \circ i$ is C^∞ as a mapping into G. Its image is in H because H is a subgroup; and by Theorem 8.5 we see that the product mapping $H \times H \to H$ is also C^∞. A similar argument shows that the mapping taking each $h \in H$ to its inverse h^{-1} is also C^∞. This completes the proof, subject to the unproved assertion concerning integral manifolds. ∎

An involutive n-plane distribution Δ on a manifold M defines what is known as a foliated structure or *foliation* on M with the maximal integral manifolds as *leaves*. Thus $M = \bigcup_{p \in M} N_p$, N_p the maximal integral manifold through p and locally the flat neighborhoods determine diffeomorphisms to $\boldsymbol{R}^m = \boldsymbol{R}^n \times \boldsymbol{R}^k$, $n + k = \dim M$, with leaves corresponding to submanifolds $\boldsymbol{R}^n \times \{(a^{n+1}, \ldots, a^m)\}$. If $n = 1$, the leaves are just integral curves of a vector field. Foliations have many interesting properties (see Haefliger [1] and Lawson [1]) since the existence of a foliation is equivalent to the existence of a rather special type (involutive) of k-plane distribution. More precisely, if a foliation is given, tangent spaces to leaves determine an involutive distribution and conversely.

Exercises

1. Show that if a distribution is involutive with respect to one choice of local basis on an open set U, the same will be true for any local basis whose domain is in U.
2. A vector field X is said to *belong* to a distribution Δ if for each $p \in M$ we have $X_p \in \Delta_p$. Show that a C^∞ distribution is involutive if and only if for every pair of C^∞-vector fields X, Y on M which belong to Δ the vector field $[X, Y]$ belongs to Δ.
3. Let N be a maximal integral manifold of a distribution Δ on M. Show that if N is closed (as a subset), then it is a regular submanifold of M.
4. Let $N \subset M$ be a submanifold, and let $X, Y \in \mathfrak{X}(M)$ such that if $p \in N$, then $X_p, Y_p \in T_p(N)$. Show that $p \in N$ then implies $[X, Y]_p \in T_p(N)$.
5. Let $F: M \to N$ be a C^∞ submersion of M onto N. Show that $F^{-1}(q)$ for all $q \in N$ are the leaves of a foliation of M.

The following problems are crucial steps in the proof of the existence of maximal integral manifolds of an involutive distribution Δ on M. Assume that $\dim \Delta = n$.

6. Let N_1, N_2 be integral manifolds of Δ with $\dim N_1 = n = \dim N_2$. If $N_1 \cap N_2 \neq \varnothing$, then $N = N_1 \cup N_2$ is an integral manifold with N_1 and N_2 as open submanifolds.
7. M may be covered by a countable collection of flat coordinate neighborhoods $\{U_i, \varphi_i\}$. If N is a connected integral manifold, $\dim N = n$, then the set $\bigcup_i (N \cap U_i)$ is a countable collection of n-slices of the U_i.
8. Continuing the notation of Exercise 6, show that $\sim$ is an equivalence relation on M: $p \sim q$ if there is a finite chain $S_1, \ldots, S_r$ of n-slices of flat neighborhoods of the collection $\{U_i, \varphi_i\}$ such that $p \in S_1$, $q \in S_r$, and $S_i \cap S_{i+1} \neq \varnothing$, $i = 1, \ldots, r-1$. [The equivalence classes can be shown to be maximal integral manifolds. The hard part is second countability.]

9 Homogeneous Spaces

In this section we consider the action of a Lie group on a manifold in a special but very important case, transitive action. Let $\theta\colon G \times M \to M$ denote such an action. Then we recall that it is transitive if for every pair $p, q \in M$, there is a $g \in G$ such that $\theta_g(p) = q$. This means that as far as properties preserved by G are concerned, any two points of the manifold are alike.

(9.1) Definition A manifold M is said to be a *homogeneous space* of the Lie group G if there is a transitive C^∞ action of G on M.

Many examples of group action have this property: $O(n)$ acts transitively on S^{n-1}, $Gl(n, \mathbf{R})$ acts transitively on $\mathbf{R}^n - (0)$ and so on; these were discussed in Section III.7. But one of the most important examples remains to be treated, since until this moment we have lacked an essential tool: Frobenius' theorem. This example, viewed first from a purely set theoretic standpoint, is the following: Let G be a group, H any subgroup, and G/H the set of left cosets. We define a left action $\lambda\colon G \times G/H \to G/H$ by $\lambda(g, xH) = gxH$; it is a left action since

(i) $\lambda(e, xH) = xH$, and
(ii) $\lambda(g_1, \lambda(g_2, xH)) = \lambda(g_1, g_2 xH) = (g_1 g_2)xH = \lambda(g_1 g_2, xH)$.

Moreover, if $\pi\colon G \to G/H$ is the natural mapping of each $g \in G$ to the coset which contains it, $\pi(g) = gH$ and if $L_g\colon G \to G$ denotes left translation, then we have the property:

(iii) $\pi \circ L_g = \lambda_g \circ \pi$ (for all $g \in G$).

The transitivity is apparent: $\lambda_{yx^{-1}}(xH) = yH$ for all $x, y \in G$.

Of course, we are far from being able to assert that when G is a Lie group, then G/H is a manifold and the mappings λ and π defined by G and H are C^∞. We did see, however, that if H is closed in G (a Lie group), then the quotient topology on G/H makes it a Hausdorff space and π an open—as well as continuous—mapping (Theorem III.7.12). We left it as an exercise to the reader to show that with this topology on G/H, λ is a continuous group action. In this section we shall go further and show that G/H is a manifold and λ is a C^∞ action. Aside from the fact that this will give us many new examples of manifolds and group action, this is important for another reason: the manifolds G/H with G acting by left translation form a universal model for all transitive actions, that is, for all homogeneous spaces.

First, consider this last statement from the set-theoretic viewpoint—without topology or differentiable structure. Let X be a set on which a group G acts transitively by the rule $\theta: G \times X \to X$. Choose, arbitrarily, a point $a \in X$ and let the isotropy subgroup (or stability group) of a be H,

$$H = \{g \in G \mid \theta_g(a) = a\}.$$

We then define a mapping $\tilde{F}: G \to X$ by $\tilde{F}(g) = \theta_g(a)$. Since θ is transitive, $\tilde{F}$ is onto; moreover for any $x \in X$, $\tilde{F}^{-1}(x) = gH$, where g is any element of G such that $\tilde{F}(g) = x$. It is then easily verified that $\tilde{F}$ induces a one-to-one onto mapping $F: G/H \to X$ by $F(gH) = \tilde{F}(g)$. For these mappings we have the relation $F \circ \pi = \tilde{F}$. Finally F carries the natural action of G on G/H, which we denoted by λ above, to the action θ, that is,

$$F \circ \lambda_g = \theta_g \circ F$$

for every $g \in G$. Thus from the set-theoretic and abstract group viewpoint, $\lambda: G \times G/H \to G/H$ is equivalent as an action to $\theta: G \times X \to X$.

Of course, it is very interesting and important to see to just what extent this still holds in the case of the transitive action of a Lie group on a manifold. Recall that by Definition III.6.17 a Lie *subgroup* H of a Lie group G is an immersed submanifold which is a Lie group with respect to the group operations of G. Since we shall use the quotient topology on G/H, we must restrict our attention to those Lie subgroups that are closed subsets if G/H is to be a Hausdorff space (Theorem III.7.12). Therefore H will be assumed to be a closed Lie subgroup. [We prove later that then H is a regular submanifold of G (compare Remark III.6.19).] A *section* V, σ on G/H will mean a continuous mapping σ of an open subset V of G/H into G, $\sigma: V \to G$, such that $\pi \circ \sigma$ is the identity on V. We then have the following basic fact.

(9.2) Theorem *Let G be a Lie group and H a closed, Lie subgroup. Then there exists a unique C^∞-manifold structure on the space G/H with the properties:* (i) *π is C^∞ and* (ii) *each $g \in G$ is in the image $\sigma(V)$ of a C^∞ section V, σ on G/H. The natural action $\lambda: G \times G/H \to G/H$ described above is a C^∞ action of G on G/H with this structure. The dimension of G/H is* dim G − dim H.

Now suppose that a Lie group G acts transitively on a manifold M, the action being given by the C^∞-mapping $\theta\colon G \times M \to M$. Using the notation above, with X replaced by M, we suppose $a \in M$ and that H is the isotropy subgroup of a. We then have the following closely related theorem to complete the picture.

(9.3) Theorem *The mapping* $\tilde{F}\colon G \to M$, *defined by* $\tilde{F}(g) = \theta(g, a)$, *is* C^∞ *and has rank equal to* $\dim M$ *everywhere on* G. *The isotropy group* H *is a closed Lie subgroup so that* G/H *is a* C^∞ *manifold. The mapping* $F\colon G/H \to M$ *defined by* $F(gH) = \tilde{F}(g)$ *is a diffeomorphism and* $F \circ \lambda_g = \theta_g \circ F$ *for every* $g \in G$.

Before proving these theorems we give some examples of their use. First consider briefly some of the spaces associated with classical geometries: $\boldsymbol{E}^n$—Euclidean space, $P^n(\boldsymbol{R})$—the space of real projective geometry, and $\boldsymbol{H}^2$—the space of plane non-Euclidean geometry. All of these were discovered and studied before Lie groups (or groups of any kind) were invented. However, in each case there is an underlying group, the group of automorphisms of the geometry; it is the group by which we can bring congruent figures into congruence. In fact each geometry studies precisely the objects and properties which are invariant under the transformations of this group acting on the space. For $\boldsymbol{E}^n$, or $\boldsymbol{R}^n$, the group consists of all isometries (rigid motions): translations, rotations, and reflections; for $P^n(\boldsymbol{R})$ it is the projective transformations; and for $\boldsymbol{H}^2$ it is the group which leaves *non-Euclidean* distances unchanged ("rigid" motions again!) In each case the group is a Lie group and in each case it is transitive. This means that the theorems above can be used as a sort of underlying unifying principle of all these geometries, a fact which was recognized by F. Klein [1] and resulted in a famous approach to geometry called the "Erlangen Program" (1872). Thus the study of any of these classical geometries can be reduced to a study of Lie groups G and their subgroups H. This point of view pervades much of modern geometry. Consider now what G and H are for the cases above.

(9.4) Example We have seen in Example III.7.6, that the group of rigid motions of $\boldsymbol{E}^n$, identified with $\boldsymbol{R}^n$, is a group G which is $O(n) \times V^n$ as a manifold, but whose group product was defined by $(A, v)(B, w) = (AB, Bv + w)$ and whose action on $\boldsymbol{R}^n$ is given by $(A, v) \cdot x = Ax + v$ (see Exercise III.7.6). Another approach is the following: we identify G with the $(n + 1) \times (n + 1)$ matrices of the form

$$g = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & v_1 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & v_n \\ \hline 0 & \cdots & 0 & 1 \end{array}\right), \qquad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in O(n),$$

and points $x = (x^1, \ldots, x^n)$ of $\boldsymbol{R}^n$ with the column vector $\tilde{x} = {}^t(x^1, \ldots, x^n, 1)$. Then $\theta(g, x) = g\tilde{x}$, the product of the matrices g and $\tilde{x}$. The subgroup H leaving the origin $x = (0, \ldots, 0)$ fixed is the set of all of these matrices for which $v_1 = \cdots = v_n = 0$; hence it is a closed Lie subgroup isomorphic to $O(n)$.

(9.5) Example The group $G = Sl(n+1, \boldsymbol{R})$ acts transitively on $P^n(\boldsymbol{R})$ as follows: let $[x] \in P^n(\boldsymbol{R})$. Then $[x]$ is an equivalence class of nonzero elements $x = (x^1, \ldots, x^{n+1})$ of $\boldsymbol{R}^{n+1}$. Given any $g \in Sl(n+1, \boldsymbol{R})$, we define $\theta(g, [x])$ by

$$\theta(g, [x]) = [gx],$$

where gx is the matrix product of g with x written as a column vector ($(n+1) \times 1$ matrix). This is a C^∞ action and is transitive; the proof is left to the exercises. The isotropy subgroup H of $[(1, 0, \ldots, 0)]$ is the set of elements (a_{ij}) of $Sl(n+1, \boldsymbol{R})$ with $a_{11} \neq 0$ and all other entries of the first column equal to zero. It is easily seen that H is a closed, Lie subgroup of G.

The non-Euclidean (or hyperbolic) plane $\boldsymbol{H}^2$ will be discussed in the last chapter. It is equivalent to $Sl(2, \boldsymbol{R})/SO(2)$ in the sense of theorem 9.3.

One of the more important uses of these ideas and of Theorem 9.2 is the relatively simple way it provides for establishing that certain sets are C^∞ manifolds in a natural way. The best illustrations are the Grassman manifolds $G(k, n)$ of k-planes through the origin in $\boldsymbol{R}^n$. It was proved in Section III.2 that these were manifolds, but the proof was quite complicated and only sketched at some points. This same result may be shown as follows. The group $Gl(n, \boldsymbol{R})$ acting in the natural manner on $\boldsymbol{R}^n$ is transitive on k-planes through the origin. This is an immediate consequence of the fact that it is transitive on n-frames: given $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ any linearly independent set of vectors, then there is a uniquely determined, nonsingular, linear transformation taking it to any second linearly independent set $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$. However, if $Gl(n, \boldsymbol{R})$ is transitive on n-frames, it is necessarily transitive on k-frames since each set of k linearly independent vectors can be completed to a basis. It follows that $Gl(n, \boldsymbol{R})$ acts transitively on the *set* $M = G(k, n)$ of k-planes through 0. If the isotropy subgroup H of some point of M, that is, a k-plane through 0, is a closed Lie subgroup, then $Gl(n, \boldsymbol{R})/H$ is a C^∞ manifold by Theorem 9.2 and is in natural one-to-one correspondence with M. Thus we may take on M the topology and C^∞ structure which makes this correspondence a diffeomorphism. However, H is such a subgroup, for the k-plane of $\boldsymbol{R}^n$ spanned by the vectors $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_k = (0, \ldots, 1, 0, \ldots, 0)$ is carried onto itself by the subgroup $H \subset Gl(n, \boldsymbol{R})$ consisting of matrices of the form

$$h = \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right),$$

where $A \in Gl(k, \boldsymbol{R})$, $B \in Gl(n-k, \boldsymbol{R})$, and C is an arbitrary $k \times (n-k)$ matrix. Therefore the Grassmann manifold $G(k, n)$ is indeed a C^∞ manifold. This method is frequently used in practice to show that some rather complicated objects can be endowed with the structure of a differentiable manifold (uniquely, according to Theorem 9.3). It may be summarized as follows:

(9.6) *If G is a Lie group and G acts on a set X transitively in such a way that the isotropy subgroup of some point a of X is a closed Lie subgroup, then there exists a (unique) C^∞ structure on X such that the action is C^∞.*

This principle as well as other results of this section are susceptible to further refinements and weakening of hypotheses (see, for example, Helgason [1, Chapter II, Sections 3 and 4]). Of course, our treatment above of $G(n, k)$ depends on Theorem 9.2, which we are now ready to prove.

Proof of Theorem 9.2 The topology on G/H is given; it is uniquely determined by the requirement that $\pi: G \to G/H$ be open and continuous. Moreover $\lambda: G \times G/H \to G/H$ is a continuous action. For let U be an open set of G/H, then we will show that $\lambda^{-1}(U)$ is open. Let W be the subset of $G \times G$ such that every pair $(g_1, g_2) \in W$ has its product $g_1 g_2$ in $\pi^{-1}(U)$, which is an open subset of G. W is open since it is the inverse image of $\pi^{-1}(U)$ under the continuous mapping $(g_1, g_2) \to g_1 g_2$. The natural mapping of $G \times G \to G \times G/H$ given by $(g_1, g_2) \to (g_1, \pi(g_2))$ is open so it carries W onto an open set which is exactly $\lambda^{-1}(U)$.

We now need to use Frobenius' theorem, which we apply to the left-invariant distribution Δ determined by $\Delta_e = T_e(H)$. As a basis Δ has any basis of left-invariant vector fields in $\mathfrak{h}$, the Lie algebra of H viewed as a subalgebra of $\mathfrak{g}$; and the integral manifolds of Δ are exactly the left cosets gH—as remarked in the previous section. It follows that there is a cubical neighborhood of e whose intersections with the cosets gH are a union of slices. To complete the proof we need a sharper result given by the following lemma, which is important in its own right.

(9.7) Lemma *If H is a connected Lie subgroup of G which is closed as a subset, then each coset gH is a closed, regular submanifold, and there is a cubical neighborhood U, φ of any $g \in G$ such that for each coset xH either $xH \cap U$ is empty or a single (connected) slice.*

Proof That H and each of its cosets is a regular submanifold is an immediate consequence of the second part of the statement, which asserts, in particular, that H and its cosets have the submanifold property (Definition III.5.1). Since each coset is a maximal integral manifold of the distribution $\Delta = \mathfrak{h}$, as we saw in the previous section, every $g \in G$ has a flat coordinate neighborhood U, φ, whose slices—determined by fixing the last $m - n$

coordinates ($n = \dim H$, $m = \dim G$)—are integral manifolds, each an open set of a coset xH of H. We must now verify that U may be taken sufficiently small that for each x, $xH \cap U$ is empty or is a *single* slice. Since Δ, integral manifolds, etc., are invariant under left translation by elements of G it is enough to check this for the special case $x = e$. If U', φ' is a flat neighborhood of e, and if $U' \cap H$ consists of a single slice, then we need only choose $U \subset U'$ small enough that $U^{-1}U \subset U'$ and so that $U, \varphi = \varphi' | U$ is still a cube in order to complete the proof. For if $x, y \in U$ are on distinct slices of U but belong to the same coset, that is, $xH = yH$, then $y^{-1}x$ and e are elements of $U' \cap H$ but lie on distinct slices (because $L_{y^{-1}}$ is a diffeomorphism and carries slices into slices). Since this contradicts our assumption about U', it cannot happen. Thus it remains to show the existence of U', φ'. We begin with an arbitrary flat neighborhood V, ψ of e, $\psi(V) = C_\delta^m(0)$, whose slices $S(a^{n+1}, \ldots, a^m) = \{q \in V | x^j(q) = a^j,\ j = n+1, \ldots, m\}$ are integral manifolds. We saw in the proof of Theorem 8.5 that the collection of distinct slices on H, that is, $V \cap H$, is countable and hence corresponds to a countable set of points $\{(a^{n+1}, \ldots, a^m)\}$ of the cube $C_\delta^{m-n}(0)$. Restricting slightly to a closed cube $\bar{V}' = \psi^{-1}(\overline{C_{\delta'}^m(0)})$, $\delta > \delta' > 0$, we may suppose this countable set is closed, for H is closed and $\bar{V}' \cap H$ is closed. Since a closed countable subset of $\mathbf{R}^{m-n}$ must contain an isolated point (Exercise 8) it follows that $H \cap V'$ contains an isolated slice. By translation invariance we may suppose this to be the slice through e. Then it is possible to choose ε', $0 < \varepsilon' < \delta'$, so that $\psi^{-1}(C_{\varepsilon'}^m(0)) = U'$ and $\varphi' = \psi | U'$ have exactly the property we have seen is needed: $H \cap U'$ is a single slice and contains the identity e. This U', φ' as we have seen enables us to complete the proof of the lemma.† ∎

Resuming the proof of Theorem 9.2, we restrict our discussion entirely to cubical neighborhoods U, φ of the type described above with $\varphi(U) = C_\varepsilon^m(0)$ and suppose that in the local coordinates $x^1, \ldots, x^n, x^{n+1}, \ldots, x^m$, the slices obtained by holding $x^{n+1}, \ldots, x^m$ fixed are the intersections of cosets gH with U. Let $A = \{q \in U \mid x^1(q) = \cdots = x^n(q) = 0\}$ and

$$\psi' \colon A \to C_\varepsilon^{m-n}(0) \subset \mathbf{R}^{m-n}$$

be defined by $\psi'(q) = (x^{n+1}(q), \ldots, x^m(q))$; A is a C^∞ submanifold of G, contained in U, and ψ' is a diffeomorphism. By our choice of U, φ we see that A meets each coset of H which intersects U in exactly one point. Therefore, π maps A homeomorphically onto an open subset V of G/H. We denote the inverse by σ; thus $\sigma\colon V \to G$ is a continuous section with $\sigma(V) = A$. Suppose that U, φ and $\tilde{U}$, $\tilde{\varphi}$ as just chosen are such that $\tilde{V} = \pi(\tilde{A})$ and $V = \pi(A)$

† As remarked earlier (compare Remark III.6.19), the conclusion follows from the much weaker hypothesis: H is an algebraic subgroup and a closed subset of G. Moreover, our assumption that H is connected is not necessary (see Exercise 9).

have common points. The set $V \cap \tilde{V}$ is open and it is not difficult to verify that the corresponding subsets $W = \sigma(V \cap \tilde{V})$ and $\tilde{W} = \tilde{\sigma}(V \cap \tilde{V})$ are diffeomorphic with respect to the natural correspondences $\tilde{\sigma} \circ \pi \colon W \to \tilde{W}$ and $\sigma \circ \pi \colon \tilde{W} \to W$ (Exercise 4). It follows that the collection of open sets $V = \pi(A)$ over all U, φ of the type above together with the homeomorphisms $\psi = \psi' \circ \sigma \colon V \to C_\varepsilon^{m-n}(0)$ determine a C^∞ structure of the type required by the conclusions of the theorem. The uniqueness follows from requirements (i) and (ii). For if we have two differentiable structures on G/H, we see that the identity is a diffeomorphism as follows: factor it locally into a section $\sigma \colon V \to G$ of the first structure followed by projection π, which is C^∞ onto the second structure. Thus the identity is a C^∞ mapping of G/H with structure one to G/H with structure two since this holds on each domain V. But the converse is also true, so the structures agree. Finally $\lambda \colon G \times G/H \to G/H$ is C^∞ since it may be written on the domain V of a section as $\lambda(g, xH) = \pi(g\sigma(x))$. This completes the proof of Theorem 9.2. ∎

We now prove the second principal result.

Proof of Theorem 9.3 $\tilde{F} \colon G \to M$ is C^∞ since $\tilde{F}(g) = \theta(g, a)$ and θ is C^∞ by assumption. From

$$\tilde{F} \circ L_g(x) = \tilde{F}(gx) = \theta_g \circ \tilde{F}(x),$$

from the chain rule, and from the fact that both L_g and θ_g are diffeomorphisms we see that the rank of $\tilde{F}$ is the same at every $g \in G$. It follows that $\tilde{F}^{-1}(a) = H$ is a closed submanifold (Theorem III.5.8) and satisfies the hypotheses of Theorem 9.2. At e we have $\tilde{F}_* \colon T_e(G) \to T_a(M)$; but each $X_e \in T_e(G)$ is the tangent vector at $t = 0$ to the curve $g(t) = \exp tX$ so that the vector $\tilde{F}_*(X_e)$ is the tangent vector to $\tilde{F}(\exp tX) = \theta(\exp tX, a)$ at a (which corresponds to $t = 0$). Since θ restricted to $g(t) = \exp tX$ is an action of R on M, then by Theorem 3.6 $\tilde{F}_*(X_e)$ is zero if and only if $\theta(\exp tX, a) = a$ for all t, that is, $\exp tX \subset H$, or $X \in T_e(H)$ the subspace of $T_e(G)$ corresponding to the subgroup H. Hence $\ker \tilde{F}_{*e} = T_e(H) = \ker \pi_{*e}$ and, as noted, $\dim \ker \tilde{F}_*$ is constant on G as is $\dim \ker \pi_*$. Since $\tilde{F}$ is onto, it follows from Theorem II.7.1 that $\dim M = \dim G - \dim H = \dim G/H$.

Now consider $F \colon G/H \to M$. Let $q \in G/H$ and V, σ a section defined on a neighborhood V of q. Since σ is C^∞ and $F \mid V = \tilde{F} \circ \sigma$, we see that F is C^∞ in a neighborhood of every point, hence C^∞ on G/H. We know that F is one-to-one and onto from set-theoretic considerations and if $\ker F_* = \{0\}$, that is, rank $F = \dim G/H = \dim M$ everywhere, then F must be a diffeomorphism. Let q be any point of G/H and suppose $q = \pi(g)$. Then using $\tilde{F} = F \circ \pi$ and the chain rule we see that $\tilde{F}_* \colon T_g(G) \to T_{\tilde{F}(g)}(M)$ is given also by $F_* \circ \pi_*$. Since $\dim \ker \tilde{F}_* = \dim \ker \pi_*$, we must have $\dim \ker F_* = 0$,

as we wished to prove. The fact that $F \circ \lambda_g = \theta_g \circ F$ was already noted; both λ_g and θ_g are diffeomorphisms, the former by Theorem 9.2 and the latter by hypothesis. This completes the proof of Theorem 9.3. ∎

Exercises

1. Verify in detail statements (i)–(iii) at the beginning of the section concerning the action of G on G/H and that if G acts transitively on a set X, then the mapping $F: G/H \to X$ defined above is in fact one-to-one, onto, and satisfies $F \circ \lambda_g = \theta_g \circ F$ as claimed.
2. Check that the isotropy subgroup $H \subset Gl(n, \boldsymbol{R})$ defined in discussing $G(n, k)$ is, in fact, closed and a Lie subgroup.
3. A sequence of n subspaces of $\boldsymbol{R}^n$, $V_n = \boldsymbol{R}^n \supset V_{n-1} \supset \cdots \supset V_1$, with $\dim V_j = j$, $j = 1, \ldots, n$, is called a *flag* of $\boldsymbol{R}^n$. Verify that the natural action of $Gl(n, \boldsymbol{R})$ on $\boldsymbol{R}^n$ is transitive on the set of flags M and use this to obtain the structure of a C^∞ manifold on M. What is $\dim M$?
4. Verify that the sets W, $\tilde{W}$ corresponding to the overlapping part $V \cap \tilde{V}$ of the domains of two sections V, σ and $\tilde{V}, \tilde{\sigma}$ as defined in the proof of Theorem 9.2 are diffeomorphic. [Note that translations on G by elements of H are diffeomorphisms and have the property that $\pi \circ R_h = \pi$. They may be used to bring a pair of corresponding points $p, \tilde{p}$ of W, $\tilde{W}$ into coincidence.]
5. Let G be a Lie group and H a closed Lie subgroup which is normal in G. Then show that G/H is a Lie group with the C^∞ structure of Theorem 9.2 and $\pi: G \to G/H$ is a Lie group homomorphism.
6. Let G_1 and G_2 be a Lie group and $F: G_1 \to G_2$ a Lie group homomorphism. Then show that the kernel of F is a closed Lie subgroup and if F is onto, then $G_2 \cong G_1/\ker F$.
7. Show that the subgroup $O(n)$ of $Gl(n, \boldsymbol{R})$ acts transitively on the Grassmann manifold $G(k, n)$ and find the isotropy subgroup of the k-plane in $\boldsymbol{R}^n$ spanned by $\mathbf{e}_1, \ldots, \mathbf{e}_k$, the first k vectors of the standard orthonormal basis (compare the remarks preceding Example III.7.4).
8. Prove that a closed countable subset of a Euclidean space $\boldsymbol{R}^k$ has an isolated point. [*Hint*: Use the Baire category theorem.]
9. Let H be a Lie group and let H_0 denote the component of e. Show that H has at most a countable number of connected components, all both open and closed and all diffeomorphic to H_0, and show that H_0 is a Lie subgroup.

Appendix Partial Proof of Theorem 4.1

Proof (Part (I) of the existence theorem (Theorem 4.1) for ordinary differential equations) We are given n functions $f^i(t, x)$ defined and of class C^r on an open subset $I_\varepsilon \times U \subset \boldsymbol{R} \times \boldsymbol{R}^n$, $I_\varepsilon = \{-\varepsilon < t < \varepsilon, \varepsilon > 0\}$. We must

show that for each $x \in U$ there is a neighborhood V and a $\delta > 0$ such that for each $a \in V$ there exist unique functions $x^i(t)$, $-\delta < t < \delta$ satisfying

$$(*) \qquad \frac{dx^i}{dt} = f^i(t, x(t))$$

and

$$(**) \qquad x^i(0) = a^i, \qquad i = 1, \ldots, n.$$

First note that if $x^i(t)$, $i = 1, \ldots, n$, are continuous functions defined for $|t| < \delta$ and they satisfy

$$x^i(t) = a^i + \int_0^t f^i(\tau, x(\tau))\, d\tau,$$

then by the fundamental theorem of calculus they are of class C^1 at least and satisfy both $(*)$ and $(**)$. From $(*)$ it then follows that they must be of class C^{r+1} at least, since their derivatives are of class C^r.

We may write this set of integral equations for $x^1(t), \ldots, x^n(t)$ as an equation in n-tuples

$$x(t) = a + \int_0^t f(t, x(t))\, dt.$$

For a given $x_0 \in U$ we choose r, $0 < r < 1$ such that $\bar{B}_{3r}(x_0) \subset U$ and an ε' satisfying $\varepsilon > \varepsilon' > 0$, so that $\bar{I}_{\varepsilon'} \subset I_\varepsilon$. Thus $f^i(t, x)$ are of class C^r, $r \geq 1$, on the compact set $\bar{I}_{\varepsilon'} \times \bar{B}_{3r}(x_0)$; therefore both the given functions f^i and their derivatives are bounded on this set. It follows that we may choose $M > 1$ such that both $M \geq \sup \|f(t, x)\|$ and $M\|x - y\| \geq \|f(t, x) - f(t, y)\|$ for all $t \in I_\varepsilon$, and $x, y \in \bar{B}_{3r}(x_0)$. The last inequality results from the mean value theorem and the continuity of the derivatives. Choose a positive δ such that $\delta < r/M^2$.

We shall prove the theorem with this δ and with $V = B_r(x_0)$—denoted B_r in what follows. Let $a \in \bar{B}_r$ and let $\mathscr{F}$ be the collection of all *continuous* maps $\varphi(t) = (\varphi^1(t), \ldots, \varphi^n(t))$ of $\bar{I}_\delta$ into $\bar{B}_{2r}(a)$ satisfying $\varphi(0) = a$. By virtue of the preceding comments it is enough to show that there is a unique member of this collection satisfying

$$\varphi = L(\varphi) = a + \int_0^t f(t, \varphi(\tau))\, d\tau$$

in order to finish the proof. This will be done by proving that $L\colon \mathscr{F} \to \mathscr{F}$ is a contracting mapping on a complete metric space and applying Theorem II.6.5; we have:

(1) $\mathscr{F}$ is a complete metric space with $d(\varphi, \psi) = \sup_{t \in \bar{I}_\delta} \|\varphi(t) - \psi(t)\|$ since this is the topology of uniform convergence of continuous functions on a compact space.

(2) If $\varphi \in \mathscr{F}$, then $L(\varphi) \in \mathscr{F}$ so that L maps $\mathscr{F}$ to $\mathscr{F}$. It is clear that $L(\varphi)$ is continuous; in fact, it is at least C^1, and when $t = 0$ the function $L(\varphi)$ has the value a. It is only necessary to check that if $|t| \le \delta$, then $\|L(\varphi)(t) - a\| \le 2r$. This results from

$$\|L(\varphi)(t) - a\| = \left\| \int_0^t f(\tau, \varphi(\tau))\, d\tau \right\| \le \int_0^t \|f(\tau, \varphi(\tau))\|\, d\tau \le M\delta < \frac{r}{M} < r.$$

(3) Finally, L is contracting. Let $\varphi, \psi \in \mathscr{F}$,

$$\begin{aligned} \|L(\varphi) - L(\psi)\| &\le \int_0^t \|f(\tau, \varphi(\tau)) - f(\tau, \psi(\tau))\|\, d\tau \\ &\le \delta M \sup_{t \in \bar{I}_\delta} \|\varphi(t) - \psi(t)\| \\ &\le \delta M\, d(\varphi, \psi) = \frac{r}{M} d(\varphi, \psi). \end{aligned}$$

But $r < 1$ and $M > 1$ so that we have

$$\|L(\varphi) - L(\psi)\| \le k\, d(\varphi, \psi), \qquad \text{where} \quad 0 < k < 1.$$

By the contracting mapping theorem there is a unique $\varphi(t)$ satisfying the conditions. This completes the proof. ∎

Notes

For those readers who wish to delve somewhat further into some of the important topics taken up in this chapter and, especially, to find complete proofs of the basic existence theorem for systems of ordinary differential equations, the existence of maximal integral manifolds for involutive distributions, and more details on homogeneous manifolds, a comment on some of the references may be helpful.

Theorem 4.1 and the material of Sections 3–5 may be found in many places. A concise and very straightforward treatment of this theorem and related material is to be found in the book of Hurewicz [1]. It is also very elegantly treated (as is Frobenius' theorem and the inverse function theorem) in Dieudonné [1, Chapter X]. Although it is well along in the book, the basic ideas can be followed without reading through all of the previous chapters. For a treatment which is specially adapted to differentiable manifolds (including local one-parameter group action) and is beautifully done see Lang [1]; the author found all of these sources very helpful. Both Dieudonné and Lang treat the subject from a very general point of view, that of Banach spaces and manifolds modeled on them. Although this may disturb some readers, it will appeal to others. In any event, it is not difficult to reduce the level of generality—the ideas and proofs are very clear. Other sources are Hirsch and Smale [1] and Arnold [1].

For Frobenius' theorem, particularly from the global point of view, as well as other topics such as Lie groups and the fundamentals of homogeneous spaces, every reader should be acquainted with the classical book by Chevalley [1], which greatly influenced all subsequent treatments. The exposition here was very much influenced by the proof of W. L. Chow. For references, a very instructive proof, and a useful generalization the reader is urged to consult Sussman [1]. The book by Warner [1] and the book of Spivak [2] should also be helpful to the

reader who wants to fill in gaps or just to read another (and more complete) treatment of the material of Sections 8 and 9. Finally, the books by Helgason [1] and Kobayashi and Nomizu [1] go much further into the ramifications of subjects treated here (especially Section 9). Many of these books have rather complete bibliographies from which the reader can search out further sources. The theorem of Cartan on closed subgroups is proved in Chevalley [1], Helgason [1], and Hochschild [1], and Adams [1].

V TENSORS AND TENSOR FIELDS ON MANIFOLDS

With some minor exceptions this chapter contains all of the basic material on tensor fields on manifolds which is used in the succeeding chapters. We limit ourselves to discussing covariant tensors since we will rarely need any other type—except vector fields; any other cases will be developed as needed.

A covariant tensor on a vector space V is simply a real-valued function $\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ of several vector variables $\mathbf{v}_1, \ldots, \mathbf{v}_r$ of V, linear in each separately (that is, multilinear). The number of variables is called the *order* of the tensor. A tensor field Φ of order r on a manifold M is an assignment to each point $p \in M$ of a tensor Φ_p on the vector space $T_p(M)$ which satisfies a suitable regularity condition (C^0, C^r, or, C^∞) as p varies on M. Sections 1 and 2 discuss the two simplest examples: $r = 1$ corresponding to functionals on a vector space and $r = 2$ corresponding to bilinear forms on a vector space. The latter includes the important case of a Riemannian metric, which is a covariant tensor field Φ of order 2 with the property that Φ_p is an inner product on $T_p(M)$ for every $p \in M$. It is the added structure given by such a tensor which enables us to measure distances, angles, volumes, and so on, on a manifold M and thus to carry over large portions of Euclidean geometry to abstract manifolds. This concept was foreshadowed in the work of Gauss for surfaces, but was discovered and expounded by Riemann for whom the metric tensor was named. Many of the consequences of Riemann's discovery are treated in Chapters VII and VIII; only one or two of them here. In particular, in Section 3 we show that a Riemannian metric gives rise to a metric $d(p, q)$ on M; and in Section 7 we show that it defines a volume element on M.

In Section 4 the basic notion of partition of unity is introduced. This is an indispensable tool used to piece together functions, mappings, forms, vector and tensor fields and other

objects, which can be defined locally on a coordinate neighborhood (that is, on an open subset of $\boldsymbol{R}^n$), to obtain a globally defined objects on M. The first application is to prove the existence of a Riemannian metric on any C^∞ manifold. Several other applications are given here and in the remainder of the chapter.

In Section 5 covariant tensor fields are discussed in more generality and in particular we treat the case of symmetric and alternating tensors—tensors which are unchanged (respectively, change sign) when two variables are interchanged. Of these two, the alternating are the most important to us since they correspond to exterior differential forms which we use frequently in the next chapters. In Section 6 we see that tensors can be added and multiplied (like any functions to the real numbers) resulting in an algebra of tensors. A slight modification of the product gives the exterior product of alternating forms. When we add to this the basic notion of derivatives of such forms, which is defined in the last section, we have all of the basic ingredients for the calculus of exterior differential forms on M. The exterior forms on M form an algebra $\Lambda(M)$ on which differentiation is a linear operator. As we shall see, this algebra plays a basic role in the geometry of manifolds.

(Sections 6–8 are used in an essential way in Chapter VI, but much of Chapter VII and parts of Chapter VIII may be read without knowledge of differential forms.)

1 Tangent Covectors

In this chapter we suppose that V is a finite-dimensional vector space over $\boldsymbol{R}$ and let V^* denote its dual space. Then V^* is the space whose elements are linear functions from V to $\boldsymbol{R}$; we shall call them *covectors*. If $\sigma \in V^*$, then $\sigma\colon V \to \boldsymbol{R}$, and for any $\mathbf{v} \in V$, we denote the value of σ on $\mathbf{v}$ by $\sigma(\mathbf{v})$ or by $\langle \mathbf{v}, \sigma \rangle$. Both notations are useful. Recall that addition and multiplication by scalars in V^* are defined by the equations

$$(\sigma_1 + \sigma_2)(\mathbf{v}) = \sigma_1(\mathbf{v}) + \sigma_2(\mathbf{v}), \qquad (\alpha\sigma)(\mathbf{v}) = \alpha(\sigma(\mathbf{v})),$$

giving the values of $\sigma_1 + \sigma_2$ and $\alpha\sigma$, $\alpha \in \boldsymbol{R}$, on an arbitrary $\mathbf{v} \in V$, the right-hand operations taking place in $\boldsymbol{R}$.

Knowledge of linear algebra is assumed, but by way of review we mention three frequently used facts.

(i) If $F_*\colon V \to W$ is a linear map of vector spaces, then it uniquely determines a dual linear map $F^*\colon W^* \to V^*$ by the prescription

$$(F^*\sigma)(\mathbf{v}) = \sigma(F_*(\mathbf{v})) \qquad \text{or} \qquad \langle \mathbf{v}, F^*(\sigma) \rangle = \langle F_*(\mathbf{v}), \sigma \rangle.$$

When F_* is injective (surjective), then F^* is surjective (injective). For composition of linear maps, we have $(G_* \circ F_*)^* = F^* \circ G^*$.

(ii) If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of V, then there exists a unique *dual basis* $\omega^1, \ldots, \omega^n$ of V^* such that $\omega^i(\mathbf{e}_j) = \delta^i_j$. (The symbol δ^i_j is zero if $i \neq j$ and $+1$ if $i = j$.)

If $\mathbf{v} \in V$, then $\omega^1(\mathbf{v}), \ldots, \omega^n(\mathbf{v})$ are exactly the components of $\mathbf{v}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. In other words $\mathbf{v} = \sum_{j=1}^n \omega^j(\mathbf{v}) e_j$. This is a consequence of the preceding definitions (see Exercise 1).

Observe that in property (i), the definition of F^* does not require the choice of a basis; therefore F^* is *naturally* or *canonically* determined by F_*. According to (ii), the vector spaces V, V^* have the same dimension, thus they must be isomorphic. There is no natural isomorphism; however, we do have the following property:

(iii) There is a natural isomorphism of V onto $(V^*)^*$ given by $\mathbf{v} \to \langle \mathbf{v}, . \rangle$, that is, $\mathbf{v}$ is mapped to the linear function on V^* whose value on any $\sigma \in V^*$ is $\langle \mathbf{v}, \sigma \rangle$. Note that $\langle \mathbf{v}, \sigma \rangle$ is linear in each variable separately (with the other fixed).

This shows that the dual of V^* is V itself, accounts for the name "dual" space, and validates the use of the symmetric notation $\langle \mathbf{v}, \sigma \rangle$ in preference to the functional notation $\sigma(\mathbf{v})$. We shall see in the next section that when further structure is assumed, for example, an inner product on V, then there is an associated natural isomorphism of V and V^*; thus V, V^*, and V^{**} can all be identified in this case. This is apt to be more a source of confusion than joy.

Covectors on Manifolds

Let M be a C^∞ manifold and assume $p \in M$. We denote by $T_p^*(M)$ the dual space to $T_p(M)$, thus $\sigma_p \in T_p^*(M)$ is a linear mapping $\sigma_p \colon T_p(M) \to \boldsymbol{R}$ and its value on $X_p \in T_p(M)$ is denoted by $\sigma_p(X_p)$ or $\langle X_p, \sigma_p \rangle$. Given a basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$, there is a uniquely determined dual basis $\omega_p^1, \ldots, \omega_p^n$ satisfying, by definition, $\omega_p^i(E_{jp}) = \delta_j^i$. The components of σ_p relative to this basis are equal to the values of σ_p on the basis vectors $E_{1p}, \ldots, E_{np}$, thus

$$\sigma_p = \sum_{i=1}^{n} \sigma_p(E_{ip})\omega_p^i;$$

this is the dual statement to property (ii) above.

Just as we defined a vector field on M, so may we define a covector field: It is a (regular) function σ, assigning to each $p \in M$ an element σ_p of $T_p^*(M)$. As with vector fields, we denote such a function by $\sigma, \lambda, \ldots$ and we denote by $\sigma_p, \lambda_p, \ldots$ its value at p, that is, the element of $T_p^*(M)$ assigned to p. If σ is a covector field and X is a vector field on an open subset U of M, then $\sigma(X)$ defines a function on U: to each $p \in U$ we assign the number $\sigma(X)(p) = \sigma_p(X_p)$. [*Note*: We often write $\sigma(X_p)$ for $\sigma_p(X_p)$ if σ is a covector field]. These remarks enable us to state the formal definition.

(1.1) Definition A C^r*-covector field* σ on M, $r \geq 0$, is a function which assigns to each $p \in M$ a covector $\sigma_p \in T_p^*(M)$ in such a manner that for any coordinate neighborhood U, φ with coordinate frames $E_1, \ldots, E_n$, the functions $\sigma(E_i)$, $i = 1, \ldots, n$, are of class C^r on U. For convenience, "covector field" will mean C^∞-covector field.

We remark that the following (apparently stronger) regularity condition could be used to replace the requirement of the definition—thereby avoiding the use of local coordinates.

(1.2) *Suppose that σ assigns to each $p \in M$ an element σ_p of $T_p^*(M)$. If σ is of class C^r, then for any C^∞-vector field X on an open subset W of M the function $\sigma(X)$ is of class C^r on W, and conversely.*

To see this we take a covering of W by coordinate neighborhoods of M (whose domains are in W); let U, φ be such a neighborhood. Then $X = \sum \alpha^i E_i$ on U, where α^i are C^∞ on U. Thus $\sigma(X) = \sum \alpha^i \sigma(E_i)$ on U and is C^r if $\sigma(E_1), \ldots, \sigma(E_n)$ are. Hence the condition just given implies $\sigma(X)$ is of class C^r on a collection of open sets covering W and so on W itself. The converse is obvious.

Note that if $E_1, \ldots, E_n$ is a field of (C^∞) frames on an open set $U \subset M$, then the dual basis at each point of U defines a field of dual bases $\omega^1, \ldots, \omega^n$ on U satisfying $\omega^i(E_j) = \delta^i_j$. We call this a field of *coframes—coordinate* coframes if $E_1, \ldots, E_n$ are coordinate frames. The $\omega^1, \ldots, \omega^n$ are of class C^∞ by the criterion just stated, and covector field σ is of class C^r if and only if for each coordinate neighborhood U, φ the components of σ relative to the coordinate coframes are functions of class C^r on U.

(1.3) Remark It is important to note that a C^r-covector field σ defines a map $\sigma\colon \mathfrak{X}(M) \to C^r(M)$ which is not only $\boldsymbol{R}$-linear but even $C^r(M)$-linear! More precisely, if $f, g \in C^r(M)$ and X and Y are vector fields on M, then

$$\sigma(fX + gY) = f\sigma(X) + g\sigma(Y),$$

for these functions are equal at each $p \in M$ (as the reader should verify).

(1.4) Example If f is a C^∞ function on M, then it defines a C^∞-covector field, which we shall denote df, by the formula

$$\langle X_p, df_p \rangle = X_p f \qquad \text{or} \qquad df_p(X_p) = X_p f.$$

For a vector field X on M this gives $df(X) = Xf$, a C^∞ function on M. This covector field df is called the *differential of f* and df_p, its value at p, the *differential of f at p*. In the case of an open set $U \subset \boldsymbol{R}^n$, we verify that it coincides with the usual notion of differential of a function in advanced calculus, and, in fact, makes it more precise. In this case the coordinates x^i of a point of U are functions on U and, by our definition, dx^i assigns to each vector X at $p \in U$ a number $X_p x^i$, its ith component in the natural basis of $\boldsymbol{R}^n$. In particular $\langle \partial/\partial x^j, dx^i \rangle = \partial x^i/\partial x^j = \delta^i_j$ so we see that $dx^1, \ldots, dx^n$ is exactly the field of coframes dual to $\partial/\partial x^1, \ldots, \partial/\partial x^n$. Now if f is a C^∞ function on U, then we may express df as a linear combination of

$dx^1, \ldots, dx^n$. We know that the coefficients in this combination, that is the components of df, are given by $df(\partial/\partial x^i) = \partial f/\partial x^i$. Thus we have

$$df = \frac{\partial f}{\partial x^1} dx^1 + \cdots + \frac{\partial f}{\partial x^n} dx^n.$$

Suppose $a \in U$ and $X_a \in T_a(\boldsymbol{R}^n)$. Then X_a has components, say, $h^1, \ldots, h^n$ and geometrically X_a is the vector from a to $a + h$. We have

$$df(X_a) = X_a f = \left(\sum h^i \frac{\partial}{\partial x^i}\right) f = \sum h^i \left(\frac{\partial f}{\partial x^j}\right)_a;$$

in particular, $dx^i(X_a) = h^i$, that is, dx^i measures the change in the ith coordinate of a point as it moves from the initial to the terminal point of X_a. The preceding formula may thus be written

$$df(X_a) = \left(\frac{\partial f}{\partial x^1}\right)_a dx^1(X_a) + \cdots + \left(\frac{\partial f}{\partial x^n}\right)_a dx^n(X_a).$$

This gives us a very good definition of the *differential of a function* f on $U \subset \boldsymbol{R}^n$: df is a field of linear functions which at any point a of the domain of f assigns to each vector X_a a number. Interpreting X_a as the displacement of the n independent variables from a, i.e., a as initial point and $a + h$ as terminal point, $df(X_a)$ approximates (linearly) the change in f between these points. [Compare this with our earlier discussion in Section II.1 of differentiability of a function; the expression above has meaning even if f is not C^∞, in fact exactly when f is differentiable in the (weak) sense of Section II.1.]

Covector Fields and Mappings

We shall give further examples of covector fields presently. First, however, we must study what happens when we map one manifold to another. Let $F: M \to N$ be a smooth mapping and suppose $p \in M$. Then, as we know, there is induced a linear map $F_*: T_p(M) \to T_{F(p)}(N)$. As we have pointed out in (i) at the beginning of this section, F_* determines a linear map $F^*: T^*_{F(p)}(N) \to T^*_p(M)$, given by the formula

(1.5) $$F^*(\sigma_{F(p)})(X_p) = \sigma_{F(p)}(F_*(X_p)).$$

In general, F_* does not map vector *fields* on M to vector fields on N. It is surprising, then, that given *any* C^r-*covector* field on N, F^* determines (uniquely) a covector field of the same class C^r on M by this formula. We state this as a theorem.

(1.6) Theorem *Let $F: M \to N$ be C^∞ and let σ be a covector field of class C^r on N. Then formula (1.5) defines a C^r-covector field on M.*

Proof If σ is the covector field on N, then for any $p \in M$, there is exactly one image point $F(p)$ by definition of mapping. It is thus clear that $F^*(\sigma)$ is defined uniquely at each point of M. Now suppose that for a point $p_0 \in M$ we take coordinate neighborhoods U, φ of p_0 and V, ψ of $F(p_0)$ so chosen that $F(U) \subset V$. If we denote the coordinate on U by $(x^1, \ldots, x^m)$ and those on V by $(y^1, \ldots, y^n)$, then we may suppose the mapping F to be given in local coordinates by

$$y^i = f^i(x^1, \ldots, x^m), \qquad i = 1, \ldots, n.$$

Let the expression for σ on V in the local coframes be written at $q \in V$ as

$$\sigma_q = \sum_{i=1}^{n} \alpha_i(q)\tilde{\omega}_q^i,$$

where $\tilde{\omega}_q^1, \ldots, \tilde{\omega}_q^n$ is the basis of $T_q^*(N)$ dual to the coordinate frames. The functions $\alpha^i(q)$ are of class C^r on V by hypothesis. Using the formula defining F^*, we see that if p is any point on U and $q = F(p)$ its image, then

$$(F^*(\sigma))_p(E_{jp}) = \sigma_{F(p)}(F_*(E_{jp})) = \sum \alpha_i(F(p))\tilde{\omega}_{F(p)}^i(F_*(E_{jp})).$$

However, we have previously in Theorem IV.1.6 obtained the formula

$$F_*(E_{jp}) = \sum_{k=1}^{n} \frac{\partial y^k}{\partial x^j} \tilde{E}_{kF(p)}, \qquad j = 1, \ldots, m,$$

the derivatives being evaluated at $(x^1(p), \ldots, x^m(p)) = \varphi(p)$. Using $\tilde{\omega}^i(\tilde{E}_j) = \delta_j^i$, we obtain

$$(F^*(\sigma))_p(E_{jp}) = \sum_{i=1}^{n} \alpha_i(F(p))\left(\frac{\partial y^i}{\partial x^j}\right)_{\varphi(p)}.$$

As p varies over U these expressions give the components of $F^*(\sigma)$ relative to $\omega^1, \ldots, \omega^m$ on U, the coframes dual to $E_1, \ldots, E_m$. They are clearly of class C^r at least, and this completes the proof. ∎

The formulas are themselves of some interest and may be used for computation, so we shall display them in a corollary.

(1.7) Corollary *Using the notation above, let $\sigma = \sum_{i=1}^n \alpha_i \tilde{\omega}^i$ on V and $F^*(\sigma) = \sum_{j=1}^m \beta_j \omega^j$ on U, where α_i and β_j are functions on V and U, respectively, and $\tilde{\omega}^i$, ω^j are the coordinate coframes. Then*

$$F^*(\tilde{\omega}^i) = \sum_{j=1}^{m} \frac{\partial y^i}{\partial x^j}\omega^j \qquad \textit{and} \qquad \beta_j = \sum_{i=1}^{n} \frac{\partial y^i}{\partial x^j}\alpha_i, \qquad i = 1, \ldots, n,$$
$$j = 1, \ldots, m.$$

The first formulas give the relation of the bases; the second those of the components. If we apply this directly to a map of an open subset of $\boldsymbol{R}^m$ into $\boldsymbol{R}^n$, these give for $F^*(dy^i)$ the formula

$$F^*(dy^i) = \sum_{j=1}^{m} \frac{\partial y^i}{\partial x^j} dx^j, \qquad i = 1, \ldots, n. \tag{1.8}$$

(1.9) Remark Let U, φ be a coordinate neighborhood on M, $\dim M = n$, and $U' = \varphi(U) \subset \boldsymbol{R}^n$. Then $dx^1, \ldots, dx^n$ on U' are dual to $\partial/\partial x^1, \ldots, \partial/\partial x^n$, and we have $\varphi_*^{-1}(\partial/\partial x^i) = E_i$ so that $\varphi_*(E_i) = \partial/\partial x^i$, $i = 1, \ldots, n$. This gives for $\varphi^*(dx^i)$

$$\langle E_i, \varphi^*(dx^j) \rangle = \langle \varphi_*(E_i), dx^j \rangle = \delta_i^j$$

This shows that the coordinate coframes $\omega^1, \ldots, \omega^n$ dual to $E_1, \ldots, E_n$ are defined by $\omega^j = \varphi^*(dx^j)$, $j = 1, \ldots, n$.

There is a potential source of confusion in notation here. The coordinates $x^1, \ldots, x^n$ can be considered as functions on U and as such have differentials dx^i defined by

$$\langle X, dx^i \rangle = X x^i,$$

the ith component of X in the coordinate frames. In particular $\langle E_j, dx^i \rangle = E_j x^i = \delta_j^i$, so that $dx^1, \ldots, dx^n$ are dual to $E_1, \ldots, E_n$ and therefore $dx^i = \omega^i$, $i = 1, \ldots, n$. Combining this with the formula above gives $dx^i = \varphi^*(dx^i)$, which is nonsense unless we are careful to distinguish x^i as (coordinate) function on $U \subset M$, on the left, from x^i as (coordinate) function on $\varphi(U) = U' \subset \boldsymbol{R}^n$, on the right (cf. Remark III.3.2).

(1.10) Example We may apply Theorem 1.6 to obtain examples of covector fields on a submanifold M of a manifold N. Let $i: M \to N$ be the inclusion map and suppose σ is a covector field on N. Then $i^*(\sigma)$ is a covector field on M called the *restriction* of σ to M. It is often denoted σ_M or simply σ. Recalling that for each $p \in M$, $T_p(M)$ is identified with a subspace of $T_p(N)$ by the isomorphism i_*, we have for $X_p \in T_p(M)$

$$\sigma_M(X_p) = (i^*\sigma)(X_p) = \sigma(i_*(X_p)) = \sigma(X_p).$$

The last equality is the identification.

As an example, let $M \subset \boldsymbol{R}^n$, and let σ be a covector field on $\boldsymbol{R}^n$, for example take $\sigma = dx^1$. Then σ restricts to a covector field σ_M on M. Note that in this example dx^1 is never zero as a covector field on $\boldsymbol{R}^n$, but on M it is zero at any point p at which the tangent hyperplane $T_p(M)$ is orthogonal to the x^1-axis.

Exercises

1. Verify properties (i)–(iii) of V, V^*, and V^{**}.
2. Let $G = Gl(n, R)$ and define n^2 covector fields σ_{ij}, $1 \le i, j \le n$, on G by $\sigma_{ij} = \sum_{k=1}^{n} y_{ik}\, dx_{kj}$, where $Y = (y_{ij})$ is the inverse of $X = (x_{ij})$. Show that these forms are invariant under $L_A\colon G \to G$, left translation by A. Further show that $\{\sigma_{ij}\}$ is a field of frames on G.
3. Let $f_1, \ldots, f_r$, $r \le n$, be C^∞ functions on an open set U of a manifold M. Prove that there are coordinates V, ψ in a neighborhood of $p \in U$ such that $f_1, \ldots, f_r$ are among the coordinate functions if and only if $df_1, \ldots, df_r$ are linearly independent at p.
4. Let $F\colon M \to N$ be a C^∞ mapping such that $F_*\colon T_p(M) \to T_{F(p)}(N)$ has kernel $= (0)$. If V, ψ is a coordinate neighborhood of $F(p)$, show that after suitable renumbering of coordinates y^i, the mapping $q \to (y^1(F(q)), \ldots, y^m(F(q))$, the first m coordinates of $F(q)$, $m = \dim M$, determines a coordinate map on a neighborhood U of p.
5. Determine the subset of $\boldsymbol{R}^2$ on which $\sigma^1 = x^1\, dx^1 + x^2\, dx^2$ and $\sigma^2 = x^2\, dx^1 + x^1\, dx^2$ are linearly independent and find a frame field dual to σ^1, σ^2 over this set.
6. Show that the restriction of $\sigma = x^1\, dx^2 - x^2\, dx^1 + x^3\, dx^4 - x^4\, dx^3$ of $\boldsymbol{R}^4$ to the sphere S^3 is never zero on S^3.
7. Show that the set $\mathscr{T}^1(M)$ of all covector fields on M, like the set $\mathfrak{X}(M)$ of all vector fields on M is a $C^\infty(M)$ module. Prove also that $\sigma \in \mathscr{T}^1(M)$ if and only if σ is a $C^\infty(M)$-linear mapping from $\mathfrak{X}(M)$ to $C^\infty(M)$.
8. Try to determine a C^∞ manifold structure on $T^*(M) = \bigcup_{p \in M} T_p^*(M)$ in such a fashion that a covector field σ on M is a C^∞ mapping from M into $T^*(M)$ and so that the natural mapping π taking each $\sigma_p \in T^*(M)$ to p is C^∞

2 Bilinear Forms. The Riemannian Metric

In the case of a vector space V over $\boldsymbol{R}$ a *bilinear form* on V is defined to be a map $\Phi\colon V \times V \to \boldsymbol{R}$ that is linear in each variable separately, i.e., for $\alpha, \beta \in \boldsymbol{R}$ and $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in V$,

$$\Phi(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2, \mathbf{w}) = \alpha\Phi(\mathbf{v}_1, \mathbf{w}) + \beta\Phi(\mathbf{v}_2, \mathbf{w}),$$

$$\Phi(\mathbf{v}, \alpha\mathbf{w}_1 + \beta\mathbf{w}_2) = \alpha\Phi(\mathbf{v}, \mathbf{w}_1) + \beta\Phi(\mathbf{v}, \mathbf{w}_2).$$

A similar definition may be made for a map Φ of a pair of vector spaces $V \times W$ over $\boldsymbol{R}$. We will not pursue this generalization at the moment except to point out that the map assigning to each pair $\mathbf{v} \in V$, $\sigma \in V^*$ a number $\langle \mathbf{v}, \sigma \rangle$, as discussed in the preceding section, is an example.

Bilinear forms on V are completely determined by their n^2 values on a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V. If $\alpha_{ij} = \Phi(\mathbf{e}_i, \mathbf{e}_j)$, $1 \le i, j \le n$, are given and if $\mathbf{v} = \sum \lambda^i \mathbf{e}_i$, $\mathbf{w} = \sum \mu^j \mathbf{e}_j$ are any pair of vectors in V, then bilinearity requires that

$$\Phi(\mathbf{v}, \mathbf{w}) = \sum_{i,\, j=1} \alpha_{ij} \lambda^i \mu^j.$$

Conversely, given the $n \times n$ matrix $A = (\alpha_{ij})$ of real numbers, the formula just given determines a bilinear form Φ. Thus there is a one-to-one correspondence between $n \times n$ matrices and bilinear forms on V once a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is chosen. The numbers α_{ij} are called the *components* of Φ relative to the basis.

We will mention some special cases which will be of interest to us. A bilinear form, or function, is called *symmetric* if $\Phi(\mathbf{v}, \mathbf{w}) = \Phi(\mathbf{w}, \mathbf{v})$, and *skew-symmetric* if $\Phi(\mathbf{v}, \mathbf{w}) = -\Phi(\mathbf{w}, \mathbf{v})$. It is easily seen that regardless of the basis chosen, these correspond, respectively, to symmetric, ${}^tA = A$, and to skew-symmetric, ${}^tA = -A$, matrices of components.

A symmetric form is called *positive definite* if $\Phi(\mathbf{v}, \mathbf{v}) \ge 0$ and if equality holds if and only if $\mathbf{v} = 0$; in this case we often call Φ an *inner product* on V. We shall be particularly interested in this case in the succeeding chapters; a vector space with an inner product is called a *Euclidean* vector space since Φ allows us to define the length of a vector, $\|\mathbf{v}\| = (\Phi(\mathbf{v}, \mathbf{v}))^{1/2}$, and the angle between vectors, as was remarked in Section I.1.

(2.1) Definition A *field Φ of C^r-bilinear forms*, $r \ge 0$, on a manifold M consists of a function assigning to each point p of M a bilinear form Φ_p on $T_p(M)$, that is, a bilinear mapping $\Phi_p \colon T_p(M) \times T_p(M) \to \boldsymbol{R}$, such that for any coordinate neighborhood U, φ the functions $\alpha_{ij} = \Phi(E_i, E_j)$, defined by Φ and the coordinate frames $E_1, \ldots, E_n$, are of class C^r. Unless otherwise stated bilinear forms will be C^∞. [To simplify notation we usually write $\Phi(X_p, Y_p)$ for $\Phi_p(X_p, Y_p)$.]

The n^2 functions $\alpha_{ij} = \Phi(E_i, E_j)$ on U are called the *components of* Φ in the coordinate neighborhood U, φ. Properties similar to those of covectors hold in this case also. As in (1.2) if Φ is a function assigning to each $p \in M$ a bilinear form, then Φ is of class C^r if and only if for every pair of vector fields X, Y on an open set U of M, the function $\Phi(X, Y)$ is C^r on U. As in Remark 1.3 we have the fact that Φ is $C^\infty(U)$-bilinear as well as $\boldsymbol{R}$-bilinear: $f \in C^\infty(U)$ implies $\Phi(fX, Y) = f\Phi(X, Y) = \Phi(X, fY)$ (Exercise 2).

Suppose $F_* \colon W \to V$ is a linear map of vector spaces and Φ is a bilinear form on V. Then the formula

$$\text{(2.2)} \qquad (F^*\Phi)(\mathbf{v}, \mathbf{w}) = \Phi(F_*(\mathbf{v}), F_*(\mathbf{w}))$$

defines a bilinear form $F^*\Phi$ on W. We have the following properties:

(i) If Φ is symmetric (skew-symmetric), then $F^*\Phi$ is symmetric (skew-symmetric).

(ii) If Φ is symmetric, positive definite, and F_* is injective, then $F^*\Phi$ is symmetric, positive definite.

In particular, this latter applies to the identity map i_* of a subspace W into V. In this case $i^*\Phi$ is just restriction of Φ to W: $(i^*\Phi)(\mathbf{v}, \mathbf{w}) = \Phi(i_*\mathbf{v}, i_*\mathbf{w}) = \Phi(\mathbf{v}, \mathbf{w})$.

Now let $F: M \to N$ be a C^∞ map and suppose that Φ is a field of bilinear forms on N. Then just as in the case of covectors this defines a field of bilinear forms $F^*\Phi$ on M by the formula for $(F^*\Phi)_p$ at every $p \in M$:

$$(F^*\Phi)(X_p, Y_p) = \Phi(F_*(X_p), F_*(Y_p)).$$

We state this in the form of a theorem.

(2.3) Theorem *Let $F: M \to N$ be a C^∞ map and Φ a bilinear form of class C^r on N. Then $F^*\Phi$ is a C^r-bilinear form on M. If Φ is symmetric (skew-symmetric), then $F^*\Phi$ is symmetric (skew-symmetric).*

Proof The proof parallels those of Theorem 1.6 and Corollary 1.7 and we analogously obtain formulas for the components of $F^*\Phi$ in terms of those of Φ. We suppose U, φ and V, ψ are coordinate neighborhoods of p and of $F(p)$ with $F(U) \subset V$. Using the notation of Theorem 1.6 and Corollary 1.7 we may write $\beta_{ij}(p) = (F^*\Phi)_p(E_{ip}, E_{jp}) = \Phi(F_*(E_{ip}), F_*(E_{jp}))$. Applying Theorem IV.1.6 as before, we have

$$\beta_{ij}(p) = \sum_{s,t=1}^{n} \frac{\partial y^s}{\partial x^i}\frac{\partial y^t}{\partial x^j}\Phi(\tilde{E}_{sF(p)}, \tilde{E}_{tF(p)}).$$

This gives the formula

$$\beta_{ij}(p) = \sum_{s,t=1} \frac{\partial y^s}{\partial x^i}\frac{\partial y^t}{\partial x^j}\alpha_{st}(F(p)), \qquad 1 \le i, j \le m, \tag{2.4}$$

for the matrix of components (β_{ij}) of $F^*\Phi$ at p in terms of the matrix (α_{st}) of Φ at $F(p)$. The functions β_{ij} thus defined are of class C^r at least on U which completes the proof, except for the statements about symmetry and skew-symmetry which are obvious consequences of (i) above. ∎

(2.5) Corollary *If F is an immersion and Φ is a positive definite, symmetric form, then $F^*\Phi$ is a positive definite, symmetric bilinear form.*

Proof All that we need to check is that $F^*\Phi$ is positive definite at each $p \in M$. Let X_p be any vector tangent to M at p. Then $F^*\Phi(X_p, X_p) = \Phi(F_*(X_p), F_*(X_p)) \ge 0$ with equality holding only if $F_*(X_p) = 0$. However, since F is an immersion, $F_*(X_p) = 0$ if and only if $X_p = 0$. ∎

(2.6) Definition A manifold M on which there is defined a field of symmetric, positive definite, bilinear forms Φ is called a *Riemannian* manifold and Φ the *Riemannian metric.* We shall assume always that Φ is of class C^∞.

The simplest example is $\boldsymbol{R}^n$ with its natural inner product

$$\Phi_a(X_a, Y_a) = \sum_{i=1}^{n} \alpha^i \beta^i \qquad \text{where} \quad X = \sum \alpha^i \frac{\partial}{\partial x^i} \quad \text{and} \quad Y = \sum \beta^i \frac{\partial}{\partial x^i}.$$

At each point we have $\Phi(\partial/\partial x^i, \partial/\partial x^j) = \delta_{ij}$ so that the matrix of components of Φ, relative to the standard basis, is constant and equals I, the identity matrix. It follows that Φ is C^∞.

Corollary 2.5 enables us to give many further examples. Any imbedded or immersed submanifold M of $\boldsymbol{R}^n$ is endowed with a Riemannian metric from $\boldsymbol{R}^n$ by virtue of the imbedding (or immersion) $F: M \to \boldsymbol{R}^n$. Thus, for example, a surface M in $\boldsymbol{R}^3$ has a Riemannian metric. The idea of the corollary in this case is very simple: If $i: M \to \boldsymbol{R}^3$ is the identity and X_p, Y_p are tangent vectors to M at p, then $i^*\Phi(X_p, Y_p) = \Phi(i_* X_p, i_* Y_p) = \Phi(X_p, Y_p)$, that is, we simply take the value of the form on X_p, Y_p considered as vectors in $\boldsymbol{R}^3$, using our standard identification of $T_p(M)$ with a subspace of $T_p(R^3)$. In particular S^2, the unit sphere of $\boldsymbol{R}^3$, has a Riemannian metric induced by the standard inner product in $\boldsymbol{R}^3$. If X_p, Y_p are tangent to S^2 at p, then $\Phi(X_p, Y_p)$ is just their inner product in $\boldsymbol{R}^3$.

Classical differential geometry deals with properties of surfaces in Euclidean space. The inner product Φ on the tangent space at each point of the surface, inherited from Euclidean space, is an essential element in the study of the geometry of the surface. It is known as the *first fundamental form* of the surface.

We terminate with a few remarks about bilinear forms on an n-dimensional vector space V. We continue the numbering from properties (i) and (ii) which precede the discussion of mappings. We define the *rank* of a form Φ on V to be the codimension of the subspace $W = \{\mathbf{v} \in V \mid \Phi(\mathbf{v}, \mathbf{w}) = 0 \ \forall\ \mathbf{w} \in V\}$, that is, rank $\Phi = \dim V - \dim W$. This concept is elaborated in the exercises. The following facts are often useful:

(iii) If Φ is a bilinear form on V, then the linear mapping $\varphi: V \to V^*$ defined by $\langle \mathbf{w}, \varphi(\mathbf{v}) \rangle = \Phi(\mathbf{w}, \mathbf{v})$ is an isomorphism onto if and only if rank $\Phi = \dim V$.

(iv) Every bilinear form Φ may be written uniquely as the sum of a symmetric and a skew-symmetric bilinear form, namely,

$$\Phi(\mathbf{v}, \mathbf{w}) = \tfrac{1}{2}[\Phi(\mathbf{v}, \mathbf{w}) + \Phi(\mathbf{w}, \mathbf{v})] + \tfrac{1}{2}[\Phi(\mathbf{v}, \mathbf{w}) - \Phi(\mathbf{w}, \mathbf{v})].$$

(v) If a skew-symmetric form Φ has a rank equal to $\dim V$, then $\dim V$ is an even number.

Exercises

1. Verify (i)–(v).
2. Show that in the C^∞ case Definition 2.1 is equivalent to the following: A field of C^∞ bilinear forms on M is a $C^\infty(M)$-bilinear mapping $\Phi: \mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M)$.
3. Show that the rank of the matrix of components $(\Phi(\mathbf{e}_i, \mathbf{e}_j))$ of a bilinear form Φ on V^n is independent of the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and is equal to the rank of Φ.
4. Let Φ be a symmetric, positive definite bilinear form on V. Show that (a) rank $\Phi = \dim V$, (b) Φ determines a natural isomorphism of V and V^*, and (c) $\Phi(v, w)^2 \geq \Phi(v, v)\Phi(w, w)$. How does (c) allow us to define to within π the angle of two vectors?
5. Show that the set $\mathscr{B}(V)$ of all bilinear forms on a vector space V is a vector space and determine its dimension. Prove that a linear map F_*: $V \to W$ of vectors spaces determines a linear map $F^*: \mathscr{B}(W) \to \mathscr{B}(V)$ with $(G_* \circ F_*)^* = F^* \circ G^*$ for composed mappings.
6. Continuing Exercise 5, let $Gl(V)$ denote the group of (linear) automorphisms of V. Show that $Gl(V)$ acts on $\mathscr{B}(V)$. Choosing a basis of V determine the action explicitly.
7. Show that the set of all elements of $Gl(V)$ leaving Φ fixed under the action of Exercise 6 form a closed subgroup. Is it a regular submanifold of $Gl(V)$?
8. Let Φ be a C^∞ field of bilinear forms and X a vector field on a manifold M. Using the one-parameter group action θ_t on M and the induced mapping θ_t^*on Φ, define a "Lie derivative" $L_X\Phi$ of Φ with respect to X.
9. Show that $\Phi(A, B) = \operatorname{tr} {}^tAB$, the trace of the transpose of A times B, defines a symmetric bilinear form on $\mathscr{M}_n(\boldsymbol{R})$, the $n \times n$ matrices over $\boldsymbol{R}$. Is it positive definite?

3 Riemannian Manifolds as Metric Spaces

The importance of the Riemannian manifold derives from the fact that it makes the tangent space at each point into a Euclidean space, with inner product defined by $\Phi(X_p, Y_p)$ $(= \Phi_p(X_p, Y_p))$. This enables us to define angles between curves, that is, the angle between their tangent vectors X_p and Y_p at their point of intersection, and lengths of curves on M, as we shall see. Thus we may study many questions concerning the geometry of these manifolds; this is a large part of the classical differential geometry of surfaces in $\boldsymbol{R}^3$.

As an example we consider the question of defining the length of a curve. Let $t \to p(t)$, $a \leq t \leq b$, be a curve of class C^1 on a Riemannian manifold M. Then its *length* L is defined to be the value of the integral

$$L = \int_a^b \left(\Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right)\right)^{1/2} dt.$$

We make several comments here: first, the integrand is a function of t alone, so a more precise notation is to denote its value at each t by $\Phi_{p(t)}(dp/dt, dp/dt)$, where $dp/dt \in T_{p(t)}(M)$ denotes the tangent vector to the curve at $p(t)$. This function is continuous by the continuity of dp/dt and Φ. Secondly, the value of the integral is independent of the parametrization. In Equation IV.(3.7), we gave the following formula for change of parameter: $dp/ds = (dp/dt)(dt/ds)$, where $t = f(s)$, $c \le s \le d$ is the new parametrization. Thus

$$\int_c^d \left(\Phi\left(\frac{dp}{ds}, \frac{dp}{ds}\right)\right)^{1/2} ds = \int_a^b \left(\Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right)\left(\frac{dt}{ds}\right)^2\right)^{1/2} \frac{ds}{dt}\, dt$$

$$= \int_a^b \left(\Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right)\right)^{1/2} dt.$$

In particular, we note that the arclength along the curve from $p(a)$ to $p(t)$, which we may denote by $s = L(t)$, gives a new parameter by the formula

$$s = L(t) = \int_a^t \left(\Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right)\right)^{1/2} dt$$

which implies

$$\frac{ds}{dt} = \left(\Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right)\right)^{1/2} \quad \text{or} \quad \left(\frac{ds}{dt}\right)^2 = \Phi\left(\frac{dp}{dt}, \frac{dp}{dt}\right).$$

Within a single coordinate neighborhood U, φ with coordinate frames $E_{1p}, \ldots, E_{np}$, we have $\Phi(E_{ip}, E_{jp}) = g_{ij}(x)$, where $\varphi(p) = x = (x^1, \ldots, x^n)$; and the curve is given by $\varphi(p(t)) = (x^1(t), \ldots, x^n(t))$, so that $L(t)$ becomes

$$s = L(t) = \int_a^t \left(\sum g_{ij}(x(t)) \frac{dx^i}{dt}\frac{dx^j}{dt}\right)^{1/2} dt.$$

This leads to the frequently used abbreviation

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}(x)\, dx^i\, dx^j$$

for the Riemannian metric in local coordinates. [This formula can be interpreted (later) in terms of multiplication of tensors.]

We note that in the case of a curve in $\boldsymbol{R}^n$ (with its standard inner product), say $p(t) = (x^1(t), \ldots, x^n(t))$ with $a \le t \le b$, then we have the familiar formula for arc length

$$L = \int_a^b \left(\sum_{i=1}^{n} (\dot{x}^i(t))^2\right)^{1/2} dt.$$

We have used $\Phi(\partial/\partial x^i, \partial/\partial x^j) = \delta_{ij}$ and $dp/dt = \sum_{i=1}^n \dot{x}^i(t)\, \partial/\partial x^i$ in our definition to obtain this.

Of course, it is not necessary to assume the curve of class C^1; weaker assumptions will do. In particular, we may suppose it is piecewise of class C^1, which we will denote D^1. We will prove the following rather useful theorem concerning Riemannian manifolds:

(3.1) Theorem *A connected Riemannian manifold is a metric space with the metric $d(p, q) =$ infimum of the lengths of curves of class D^1 from p to q. Its metric space topology and manifold topology agree.*

Proof Since M is arcwise connected, $d(p, q)$ is defined; and from the definition it is immediate that $d(p, q)$ is symmetric and nonnegative. It is also very easy to check that the triangle inequality is satisfied if we use the fact that a curve from p_1 to p_2 and a curve from p_2 to p_3 may be joined to give a curve from p_1 to p_3 whose length is the sum of the lengths of the two curves which are thus joined. It remains to show that $d(p, q) = 0$ implies $p = q$.

In order to complete the proof we obtain some inequalities. In all that follows let p be an arbitrary point of M, U, φ a coordinate neighborhood which has the property that $\varphi(p) = (0, \ldots, 0)$, and $a > 0$ a fixed real number with the property that $\varphi(U) \supset \bar{B}_a(0)$, the closure of the open ball of radius a and center at the origin of $\mathbf{R}^n$. We let $x^1, \ldots, x^n$ denote the local coordinates and $g_{ij}(x)$ the components of the metric tensor Φ as functions of these coordinates. Since these n^2 functions are C^∞ in their dependence on the coordinates and are the coefficients of a positive definite, symmetric matrix for each value of x in $\varphi(U)$, then on the compact subset K_r, $r \leq a$, of $\mathbf{R}^n \times \mathbf{R}^n$ defined by $K_r = \{(x, \alpha) \mid \|x\| \leq r, \|\alpha\| = 1\}$, the expression $(\sum_{i,j=1}^n g_{ij}(x)\alpha^i\alpha^j)^{1/2}$ assumes a maximum value M_r and a minimum value $m_r > 0$. In fact if m, M denote the minimum and maximum corresponding to $r = a$, we have the inequalities

$$0 < m \leq m_r \leq \left(\sum_{i,j=1}^n g_{ij}(x)\alpha^i\alpha^j \right)^{1/2} \leq M_r \leq M.$$

Moreover, if $(\beta^1, \ldots, \beta^n)$ are any n real numbers such that $(\sum_{i=1}^n (\beta^i)^2)^{1/2} = b \neq 0$, then replacing each α^i above by β^i/b and multiplying the inequalities by b yields:

$$0 \leq mb \leq m_r b \leq \left(\sum_{i,j=1}^n g_{ij}(x)\beta^i\beta^j \right)^{1/2} \leq M_r b \leq Mb$$

for every $x \in \bar{B}_r(0)$. Now we shall make the assumption that if x, y are any points of $\mathbf{R}^n$ with its standard Riemannian metric (as defined above), then the infimum of the lengths of all D^1 curves in $\mathbf{R}^n$ from x to y is exactly the length of the line segment $\overline{xy}$, in other words, it is $\|y - x\|$ the Euclidean distance from x to y as defined in Section I.1 (see Exercises 5 and 6 for the

method of proof). Let $p(t)$, $a \le t \le b$, be a D^1 curve lying in $\varphi^{-1}(\bar{B}_r(0)) \subset U$ which runs from $p = p(a)$ to $q = p(b)$ and let

$$L = \int_a^b \left[\sum_{i,j=1}^n g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t) \right]^{1/2} dt$$

denote its length. The last set of inequalities above and the assumption on $\boldsymbol{R}^n$ imply that for $p \neq q$

$$0 < m\|\varphi(q)\| \le m_r\|\varphi(q)\| \le L \le M_r \int_a^b \left[\sum_{i=1}^n (\dot{x}^i)^2 \right]^{1/2} dt$$
$$\le M \int_a^b \left[\sum_{i=1}^n (\dot{x}^i)^2 \right]^{1/2} dt.$$

We first use these inequalities to complete the proof that $d(p, q)$ is a metric on M. Let q' be any point of M distinct from p. Then for some r, $0 < r \le a$, q' lies outside of $\varphi^{-1}(B_r(0)) \subset U$. Let $p(t)$, $0 \le t \le c$, be a curve of class D^1 which goes from $p = p(0)$ to $q' = p(c)$ and let L' be its length. There is a first point $q = p(b)$ on the curve which is outside $\varphi^{-1}(B_r(0))$, that is, such that $p(t)$ lies inside the neighborhood $\varphi^{-1}(B_r(0))$ for $0 \le t < b$, but $q = p(b)$ does not; q is the first point of the curve with $\|\varphi(q)\| = r$. If L denotes the length of the curve $p(t)$, $0 \le t \le b$, then $L \le L'$. From this it follows that $L' \ge L \ge mr$ and, since the curve was arbitrarily chosen, that $d(p, q) \ge mr$. This means that if $q' \neq p$, then $d(p, q') \neq 0$, so that $d(p, q)$ is a metric as claimed.

In order to show the equivalence of the metric and the manifold topologies on M, it is enough to compare the neighborhood systems at an arbitrary point p of M; in fact for the manifold topology we need only consider the neighborhoods lying inside a single coordinate neighborhood U, φ as above. Thus we must show that each neighborhood $V_r = \varphi^{-1}(B_r(0)) \subset U$ of the point p contains an ε-ball, $S_\varepsilon(p) = \{q \in M \mid d(p, q) < \varepsilon\}$, of the metric topology, and conversely. This will follow from the inequalities we have obtained. In the last set of inequalities, if $r \le a$ and $q \in \varphi^{-1}(B_r(0))$, then $\|\varphi(q)\|$ is the Euclidean distance $d_{\mathrm{E}}(p, q)$ of p from q, i.e., the distance in $\boldsymbol{R}^n$ of $\varphi(p) = 0$ from $\varphi(q)$. Thus $V_r = \{q \in U \mid d_{\mathrm{E}}(p, q) < r\}$. The left side of the inequality can be written as

$$0 < m d_{\mathrm{E}}(p, q) \le m_r d_{\mathrm{E}}(p, q) \le L.$$

Since $d(p, q)$ is the infimum of such L, we see that $V_r \subset S_\varepsilon(p)$ for $r < \varepsilon/m$.

On the other hand, given V_r, $r \le a$, and a D' curve in $\bar{V}_r = \varphi^{-1}(\bar{B}_r(0))$, as described above, the integral on the right-hand side of the inequality is its Euclidean length L_{E}, i.e., the length of $\varphi(p(t))$, $a \le t \le b$. Thus the right-hand side of the inequality may be written as

$$L < M_r L_{\mathrm{E}} \le M L_{\mathrm{E}}.$$

According to Exercises 5 and 6, $d_E(p, q)$ is the infimum of the lengths of such curves from p to q, just as $d(p, q)$ is the infimum of L for such curves. Hence, $S_\varepsilon(p) \subset V_r$ if $\varepsilon < Mr$. Thus the $S_\varepsilon(p)$ and V_r form equivalent neighborhood systems at p. This completes the proof of the theorem except for the unproved assertion about $\boldsymbol{R}^n$ (essentially this theorem itself in $\boldsymbol{R}^n$), which is left to the exercises. ∎

As we have mentioned, the existence of a Riemannian metric on a manifold provides an important ingredient to the study of manifolds from a *geometric* point of view, allowing us to introduce on such spaces many concepts of Euclidean geometry such as distances, angles between curves, areas, volumes, and—less obviously—straight lines, or *geodesics*. For one way of characterizing a straight line in Euclidean space is that the length of any segment $\overline{pq}$ on it is exactly the distance $d(p, q)$ between its end points—which implies that it is also the shortest curve between any two of its points (by Exercises 5 and 6 again). We can, using the metric just introduced, ask whether there exist curves on a Riemannian manifold which have this property. The answer, with some qualifications, is yes; and the class of curves (geodesics) thus isolated has both similarities to and fascinating differences from straight lines in Euclidean geometry. For example, if S^2 is the unit sphere in $\boldsymbol{R}^3$, with the induced Riemannian metric, then great circles are the geodesics: they indeed realize the distance (and are the shortest curves) between any two of their points which lie on the same semicircle. Note that these geodesics are *closed curves* in marked contrast to straight lines in Euclidean space.

Two Riemannian manifolds M_1 and M_2 (with Riemannian metrics Φ_1 and Φ_2) are said to be *isometric* if there exists a diffeomorphism $F: M_1 \to M_2$ such that $F^*\Phi_2 = \Phi_1$. Clearly such an isometry is also an isometry of M_1 and M_2 as metric spaces, that is, $d_2(F(p), F(q)) = d_1(p, q)$ in the metrics defined above. It is true, but not easy to prove, that a converse to this statement holds (see Kobayashi and Nomizu [1, Theorem 3.10, p. 169]).

The geometry, including geodesics, lengths of curves, areas, and so forth, depends very much on the Riemannian metric of M. For example, the sphere S^2 as an abstract manifold is diffeomorphic to many surfaces in $\boldsymbol{R}^3$, of which the unit sphere is only one possibility, another being a standard ellipsoid E in $\boldsymbol{R}^3$:

$$E = \left\{ x \in \boldsymbol{R}^3 \;\middle|\; \frac{(x^1)^2}{a^2} + \frac{(x^2)^2}{b^2} + \frac{(x^3)^2}{c^2} = 1 \right\}.$$

Here the geodesics of the induced Riemannian metric, unlike great circles on the unit sphere S^2, are not closed curves in general (see Hilbert and Cohn-Vossen [1, p. 222–4]). Thus E and S^2 are diffeomorphic but not isometric, that is, they are equivalent as differentiable manifolds but not as Riemannian

manifolds. An important question is to decide whether or not two given Riemannian manifolds are, in fact, isometric; and if so, in how many ways. One of the questions which led Gauss to some of his great discoveries may have been a very practical one: Is there any isometry possible between a portion of the surface of a sphere (the earth) with the metric mentioned above and a portion of the Euclidean plane with its standard metric? Or equivalently, can we construct a map of some part of the earth's surface which does not distort distances and/or angles? We shall come back to this question in a later chapter.

Exercises

1. Using spherical coordinates (θ, φ) on the unit sphere $\rho = 1$ in $\boldsymbol{R}^3$, determine the components (g_{ij}) of the Riemannian metric on the domain of the coordinates ($U = S^2$ minus the north and south poles).
2. Similarly, find g_{ij} for $T^2 = S^1 \times S^1$ using coordinates (θ, φ) and the imbedding
$$(\theta, \varphi) \to ((a + b \cos \varphi) \cos \theta, (a + b \cos \varphi) \sin \theta, b \sin \varphi)$$
in $\boldsymbol{R}^3$ given by rotating a circle of radius b, center at $(a, 0, 0)$, $a > b$, around the x^3-axis.
3. Show that to each vector field X on a Riemannian manifold there corresponds a uniquely determined covector field σ_X by (iii) of the previous section. Show that this is actually an $\boldsymbol{R}$-linear map. Is it $C^\infty(M)$-linear? Use your arguments to show that if $f \in C^\infty(M)$, then there is a unique C^∞ vector field, grad f, which corresponds via the metric to $\sigma = df$.
4. Using the results of Exercise 8 of Section 2, show that θ_t is a 1-parameter group of isometries of a Riemannian manifold M if and only if the Lie derivative of the Riemannian metric Φ with respect to the infinitesimal generator of θ_t is zero, $L_X \Phi = 0$.
5. Let $x(t)$, $0 \leq t \leq 1$, be a curve of class D^1 in $\boldsymbol{R}^n$ from $x(0) = (0, 0, \ldots, 0)$ to $x(1) = (a^1, \ldots, a^n)$. Assume, for simplicity, that $\|x(t)\| > 0$ for $t > 0$ and write $x(t) = \lambda(t)\mathbf{u}(t)$, where $\lambda(t) = \|x(t)\|$ and $\mathbf{u}(t)$ is a unit vector. Show that $\|\dot{x}(t)\|^2 = (\dot{\lambda}(t))^2 + (\lambda(t))^2\|\dot{\mathbf{u}}(t)\|^2$, and use this to prove that the length of the curve is at least $\|x(1) - x(0)\|$, the distance from the origin to $a = (a^1, \ldots, a^n)$.
6. Show that the simplifying assumption that $\|x(t)\| > 0$ for $t > 0$ in Exercise 5 may be removed by considering only the portion of the curve outside a small sphere around the origin, whose radius we then let tend to zero. Use these results to establish that in $\boldsymbol{R}^n$ the infimum of the length of curves of class D^1 joining two points x and y is $\|x - y\|$ so that the metric defined on $\boldsymbol{R}^n$ by Theorem 3.1 and the standard Riemannian structure is the usual one.

7. Use the Riemannian metric to define the cosine of the angle between intersecting curves. Apply this to the coordinate curves in Exercises 1 and 2.

4 Partitions of Unity

We have mentioned, but not proved, that there is no nonvanishing C^∞-vector field on S^2. It follows from the exercises at the end of the last section that the same is true for covector fields on S^2. In view of this nonexistence, it might occur to ask whether on an arbitrary manifold M it is possible to define a C^∞ positive definite, bilinear form, that is, is every manifold Riemannian? This question and a number of others may be answered using the notion of a partition of unity. Before discussing this concept we need some preliminary definitions and lemmas.

A covering $\{A_\alpha\}$ of a manifold M by subsets is said to be *locally finite* if each $p \in M$ has a neighborhood U which intersects only a finite number of sets A_α. If $\{A_\alpha\}$ and $\{B_\beta\}$ are coverings of M, then $\{B_\beta\}$ is called a *refinement* of $\{A_\alpha\}$ if each $B_\beta \subset A_\alpha$ for some α. In these definitions we do not suppose the sets to be open. Any manifold M is locally compact since it is locally Euclidean; it is also *σ-compact*, which means that it is the union of a countable number of compact sets. This follows from the local compactness and the existence of a countable basis $P_1, P_2, \ldots$ such that each $\bar{P}_i$ is compact. A space with the property that every open covering has a locally finite refinement is called *paracompact*; it is a standard result of general topology that a locally compact Hausdorff space with a countable basis is paracompact. We will prove a version of this adapted to our needs.

(4.1) Lemma *Let $\{A_\alpha\}$ be any covering of a manifold M of dimension n by open sets. Then there exists a countable, locally finite refinement $\{U_i, \varphi_i\}$ consisting of coordinate neighborhoods with $\varphi_i(U_i) = B_3^n(0)$ for all $i = 1, 2, 3, \ldots$ and such that $V_i = \varphi_i^{-1}(B_1^n(0)) \subset U_i$ also cover M.*

Proof Since the lemma is trivial otherwise, we assume that M is not compact. We begin with the countable basis of open sets $\{P_i\}$, $\bar{P}_i$ compact, which we mentioned above. Define a sequence of compact sets $K_1, K_2, \ldots$ as follows: $K_1 = \bar{P}_1$ and, assuming $K_1, \ldots K_i$ defined, let r be the first integer such that $r > i$ and $K_i \subset \bigcup_{j=1}^{r} P_j$. Define K_{i+1} by

$$K_{i+1} = \bar{P}_1 \cup \bar{P}_2 \cup \cdots \cup \bar{P}_r = \overline{P_1 \cup \cdots \cup P_r}\,.$$

Denote by $\mathring{K}_{i+1}$ the interior of K_{i+1}; it contains K_i. Let $K_{-1} = K_0 = \varnothing$ and for each $i = 0, 1, 2, \ldots$, we consider the open set $(\mathring{K}_{i+2} - K_{i-1}) \cap A_\alpha$. Around each p in this set choose a coordinate neighborhood $U_{p,\alpha}, \varphi_{p,\alpha}$ lying inside the set and such that $\varphi_{p,\alpha}(p) = 0$ and $\varphi_{p,\alpha}(U_{p,\alpha}) = B_3^n(0)$. Take

$V_{p,\alpha} = \varphi_{p,\alpha}^{-1}(B_1^n(0))$ and note that these are also interior to $(\mathring{K}_{i+2} - K_{i-1}) \cap A_\alpha$. Moreover allowing p, α to vary, a finite number of the collection of $V_{p,\alpha}$ covers $K_{i+1} - \mathring{K}_i$, a closed compact set. Denote these by $V_{i,k}$ with k labeling the sets in this finite collection. For each $i = 1, 2, \ldots$ the index k takes on just a finite number of values; thus the collection $V_{i,k}$ is denumerable. We renumber them as $V_1, V_2, \ldots$, and denote by $U_1, \varphi_1, U_2, \varphi_2, \ldots$, the corresponding coordinate neighborhoods containing them. These satisfy the requirements of the conclusion; in fact for each $p \in M$ there is an index i such that $p \in \mathring{K}_i$ but from the definition of U_j, V_j it is clear that only a finite number of these neighborhoods meet $\mathring{K}_i$. Therefore $\{U_i\}$ and also $\{V_i\}$ are locally finite coverings refining the covering $\{A_\alpha\}$. ∎

(4.2) Remark It is clear that it would be possible to replace the spherical neighborhoods $B_r^n(0)$ by cubical neighborhoods $C_r^n(0)$ in the lemma.

We shall call the refinement U_i, V_i, φ_i obtained in this lemma a *regular covering by spherical* (or, when appropriate, *cubical*) *coordinate neighborhoods* subordinate to the open covering $\{A_\alpha\}$.

Recall that the *support* of a function f on a manifold M is the set $\operatorname{supp}(f) = \overline{\{x \in M \mid f(x) \neq 0\}}$, the closure of the set on which f does not vanish.

(4.3) Definition A C^∞ *partition of unity on* M is a collection of C^∞ functions $\{f_\gamma\}$ defined on M with the following properties:

(1) $f_\gamma \geq 0$ on M,
(2) $\{\operatorname{supp}(f_\gamma)\}$ form a locally finite covering of M, and
(3) $\sum_\gamma f_\gamma(x) = 1$ for every $x \in M$.

Note that by virtue of (2) the sum is a well-defined C^∞ function on M since each point has a neighborhood on which only a finite number of the f_γ's are different from zero. A partition of unity is said to be *subordinate* to an open covering $\{A_\alpha\}$ of M if for each γ, there is an A_α such that $\operatorname{supp}(f_\gamma) \subset A_\alpha$.

(4.4) Theorem *Associated to each regular covering $\{U_i, V_i, \varphi_i\}$ of M there is a partition of unity $\{f_i\}$ such that $f_i > 0$ on $V_i = \varphi_i^{-1}(B_1(0))$ and* $\operatorname{supp} f_i \subset \varphi_i^{-1}(\bar{B}_2(0))$. *In particular, every open covering $\{A_\alpha\}$ has a partition of unity which is subordinate to it.*

Proof Exactly as in part (a) of the proof of Theorem II.5.1, we see that there is, for each i, a nonnegative C^∞ function $\tilde{g}(x)$ on $\boldsymbol{R}^n$ which is identically one on $\bar{B}_1^n(0)$ and zero outside $B_2^n(0)$. Clearly g_i, defined by $g_i = \tilde{g} \circ \varphi_i$ on U_i and $g_i = 0$ on $M - U_i$, is C^∞ on M. It has its support in $\varphi_i^{-1}(\bar{B}_2^n(0))$, is $+1$

on $\bar{V}_i$, and is never negative. From these facts and the fact that $\{V_i\}$ is a locally finite covering of M we see that

$$f_i = \frac{g_i}{\sum_j g_j}, \qquad i = 1, 2, \ldots,$$

are functions with the desired properties. ∎

Some Applications of the Partition of Unity

We shall give several applications which illustrate the utility of this concept. The first answers the question raised at the beginning of the first paragraph in this section.

(4.5) Theorem *It is possible to define a C^∞ Riemannian metric on every C^∞ manifold.*

Proof Let $\{U_i, V_i, \varphi_i\}$ be a regular covering of M and f_i an associated C^∞ partition of unity as defined above. Then $\varphi_i \colon U_i \to B_3^n(0)$ being a diffeomorphism, the bilinear form $\Phi_i = \varphi_i^*\Psi$, Ψ the usual Euclidean inner product on $\boldsymbol{R}^n$, defines a Riemannian metric on U_i (Corollary 2.5). Since $f_i > 0$ on V_i, $f_i \Phi_i$ is a Riemannian metric tensor on V_i, is symmetric on U_i, and is zero outside $\varphi_i^{-1}(\bar{B}_2^n(0))$. Hence it may be extended to a C^∞-symmetric bilinear form on all of M which vanishes outside $\varphi_i^{-1}(\bar{B}_2^n(0))$ but is positive definite at every point of V_i. It is easy to check that the sum of symmetric forms is symmetric, therefore $\Phi = \sum f_i \Phi_i$, defined precisely by

$$\Phi_p(X_p, Y_p) = \sum_{i=1}^{\infty} f_i(p)\Phi_i(X_p, Y_p), \qquad p \in M,$$

is symmetric. We have denoted by $f_i\Phi_i$ its extension to all of M, and we must remember that the summation makes sense since in a neighborhood of each $p \in M$ all but a finite number of terms are zero. However, Φ is also positive definite. For every i, $f_i \geq 0$ and each $p \in M$ is contained in at least one V_j. Then $f_j(p) > 0$, and therefore

$$0 = \Phi_p(X_p, X_p) = \sum f_i(p)\Phi_i(X_p, X_p)$$

implies that $\Phi_j(X_p, X_p) = 0$. This means

$$0 = \varphi_j^*\Psi(X_p, X_p) = \Psi(\varphi_{j*}(X_p), \varphi_{j*}(X_p)).$$

But since Ψ is positive definite and φ is a diffeomorphism, this implies $X_p = 0$. This completes the proof. ∎

As a second application we consider the following question. Let M be a C^∞ manifold. Then is M diffeomorphic to a submanifold of Euclidean space $\boldsymbol{R}^N$ of some sufficiently high dimension N? This is a rather difficult question,

particularly if we modify it slightly so as to leave the choice of N less arbitrary. For example, is every surface, that is, every two-dimensional manifold M, imbeddable as a submanifold of $\boldsymbol{R}^3$? [The answer is no; it is known that this is not always possible even if the surface is compact.] We shall give a partial answer to the question as first posed.

(4.6) Theorem *Any compact C^∞ manifold M admits a C^∞ imbedding as a submanifold of $\boldsymbol{R}^N$ for sufficiently large N.*

Proof We let $\{U_i, V_i, \varphi_i\}$ be a *finite* regular covering of M; there exists such because of the compactness. We have defined the associated partition of unity $\{f_i\}$ using functions $\{g_i\}$, where $g_i = 1$ on V_i, and we shall use here these C^∞ functions $\{g_i\}$ on M rather than the (normalized) $\{f_i\}$. Let $\varphi_i \colon U_i \to B_3^n(0)$ be the coordinate map. Then the mapping $g_i \varphi_i \colon U_i \to B_3^n(0)$—mapping $p \in U_i$ to $g_i(p)\varphi_i(p) = (g_i(p)x^1(p), \ldots, g_i(p)x^n(p))$ in $\boldsymbol{R}^n$—is a C^∞ map on U_i, taking everything outside $\varphi_i^{-1}(B_2^n(0))$ onto the origin, but agreeing with φ_i on V_i. It may be extended to a C^∞ mapping of M into $B_3^n(0)$ by letting it map all of $M - U_i$ onto the origin. When we write $g_i \varphi_i$, we will mean this extension; on V_i it is a diffeomorphism to $B_1^n(0)$ so its Jacobian matrix has rank $n = \dim M$ there.

We suppose that $i = 1, \ldots, k$ is the range of indices in our finite regular covering and let $N = (n + 1)k$. Define

$$F \colon M \to \boldsymbol{R}^N \to \underbrace{\boldsymbol{R}^n \times \cdots \times \boldsymbol{R}^n}_{k} \times \underbrace{\boldsymbol{R} \times \cdots \times \boldsymbol{R}}_{k}$$

by

$$F(p) = (g_1(p)\varphi_1(p); \ldots ; g_k(p)\varphi_k(p); g_1(p), \ldots, g_k(p)).$$

Then F is clearly C^∞ on M, and in any local coordinates on M the $N \times n$ Jacobian of F breaks up into k blocks of size $n \times n$ followed by a $k \times n$ matrix so its rank is at most n. However, $p \in M$ implies $p \in V_i$ for some i, and, on V_i, $g_i \equiv 1$, so $g_i \varphi_i \equiv \varphi_i$ and the matrix has rank n. Thus $F \colon M \to \boldsymbol{R}^N$ is a C^∞ immersion. If it is one-to-one the proof is finished since M is compact and Theorem III.5.7 applies. Suppose $F(p) = F(q)$. Then $g_i(p) = g_i(q)$, $i = 1, \ldots, k$. This implies that $g_i(p)\varphi_i(p) = g_i(q)\varphi_i(q)$; but since $g_i(p) \neq 0$ for some i, this means $\varphi_i(p) = \varphi_i(q)$ for that i and since φ_i is one-to-one, we see that $p = q$. Thus F is one-to-one, completing the proof. ∎

We remark that it is an obvious disadvantage of this theorem that N may be much larger than we would like it; in fact we have no way of giving an effective bound on it from this proof. For example, we know that it takes at least two coordinate neighborhoods to cover S^2 (using stereographic projections from the north and south poles) and hence $k = 2, n = 2$ so that $N = 6$, implying that S^2 may be imbedded in $\boldsymbol{R}^6$. This is obviously not the best

possible! Since a sphere with handles (compare Section I.4) may require more than two coordinate neighborhoods to cover it, the value of N would increase accordingly.

Another defect of the theorem is that it only applies to a *compact* manifold and although such manifolds are important, it would be very nice to know that *every* manifold may be considered as a submanifold of $\boldsymbol{R}^N$ for some N. Then our intuitive geometric concepts derived from the classical study of curves and surfaces in $\boldsymbol{R}^3$ could be seen to carry over to arbitrary manifolds; in particular, the concept of tangent space $T_p(M)$ is given intuitive content just as in Example IV.1.10. Clearly this question has great interest in manifold theory. The following theorem was proved by Whitney [1] in 1936 in a paper which is one of the landmarks in the study of differentiable manifolds. It is known as the Whitney imbedding theorem.

(4.7) Theorem *Any differentiable manifold M of dimension n may be imbedded differentiably as a closed submanifold of $\boldsymbol{R}^{2n+1}$.*

The proof has since been simplified and appears in many recent texts, for example, Milnor [2], Sternberg [1], and Auslander and MacKenzie [1] and Hirsch [1]. (Whitney later, by more difficult methods, replaced $\boldsymbol{R}^{2n+1}$ by $\boldsymbol{R}^{2n}$.)

Our final example of the way in which the ideas of this section may be used will be to prove the following "smoothing" theorem:

(4.8) Theorem *Let M be a C^∞ manifold and A a compact subset of M, possibly empty. If g is a continuous function on M which is C^∞ on A and ε is a positive continuous function on M, then there exists a C^∞ function h on M such that $g(p) = h(p)$ for every $p \in A$ and $|g(p) - h(p)| < \varepsilon(p)$ on all of M.*

In order to prove this we shall need a similar theorem for the case of a closed n-ball in $\boldsymbol{R}^n$. For convenience we choose the following one (see Dieudonné [1] for a proof).

(4.9) Lemma (Weierstrass approximation theorem) *Let f be a continuous function on a closed n-ball $\bar{B}^n$ of $\boldsymbol{R}^n$ and let $\varepsilon > 0$. Then there is a polynomial function p on $\boldsymbol{R}^n$ such that $|f(x) - p(x)| < \varepsilon$ on $\bar{B}^n$.*

Another similar but easier approximation lemma for $\boldsymbol{R}^n$ which would serve equally well here is given in Section VI.8 (Exercise 2). We now proceed with the proof of the theorem.

Proof By definition "g is C^∞ on A" means that there is a C^∞ extension g^* of $g|A$ which is C^∞ on an open set W' containing A. Since $g - g^* = 0$ on A, which is compact, we can find an open set W, $A \subset W \subset W'$ such that

$|g(p) - g^*(p)| < \varepsilon(p)$ for $p \in W$. To do this, first find a finite covering of A by neighborhoods W_i such that $\bar{W}_i \subset W'$ is compact. Then $K = \bigcup \bar{W}_i \subset W$ is compact and we let $\varepsilon' = \inf_{p \in K} \varepsilon(p) > 0$. Finally, let W be the open subset of $\bigcup W_i$ on which $|g - g^*| < \varepsilon'$.

Next take a regular covering $\{U_i, \varphi_i\}$ which is a refinement of covering $\{W, M - A\}$ of M. On each U_i we define a C^∞ function h_i as follows: If $U_i \subset W$, then let $h_i = g^*|U_i$. Otherwise, we use the Weierstrass approximation theorem to obtain a polynomial function p_i on $\boldsymbol{R}^n$ which approximates $\hat{g} = g \circ \varphi_i^{-1}$ on $\bar{B}_2^n(0) \subset \varphi(U_i)$ to within $\varepsilon_i' = \inf \varepsilon(p)$ on $p \in \varphi_i^{-1}(\bar{B}_2^n(0))$, and then define $h_i = p_i \circ \varphi_i$. Thus on each U_i we have defined a C^∞ function h_i which satisfies $|h_i - g| < \varepsilon(p)$ on $\varphi_i^{-1}(\bar{B}_2^n(0))$. Now let $\{f_i\}$ be the C^∞ partition of unity subordinate to our regular covering. For each i we can interpret $f_i h_i$ as a C^∞ function on M by taking it to be 0 outside supp f_i, which lies in $\varphi_i^{-1}(\bar{B}_2^n(0))$. Thus $h = \sum f_i h_i$ is defined and C^∞ on all of M. Moreover, since $\sum f_i \equiv 1$,

$$|g - h| = |\sum f_i g - \sum f_i h_i| \leq \sum |f_i(g - h_i)|.$$

Now if $p \notin \varphi_i^{-1}(\bar{B}_2^n(0))$, then $f_i(p)(g(p) - h_i(p)) = 0$ and if $p \in \varphi_i^{-1}(\bar{B}_2^n(0))$, then $|g(p) - h_i(p)| < \varepsilon(p)$ by our choice of h_i. Since $\sum f_i(p)$ over the finite set of i for which $p \in \varphi_i^{-1}(\bar{B}_2^n(0))$ is 1 and $f_i \geq 0$ everywhere, we see that $|g(p) - h(p)| < \varepsilon(p)$ for all $p \in M$. ∎

(4.10) Remark Techniques of this type are very important in bridging the gap between the applications to manifolds of topology—where the data are usually continuous—and of calculus concepts such as rank of a mapping. By using the fact that a manifold is σ-compact, for example, and reverting to the use of local coordinates, it is possible to prove by methods of this section such statements as the following (compare Steenrod [1, p. 25]):

Let X be a vector field which is continuous on all of M and C^∞ on the closed subset A. Then X may be approximated arbitrarily closely by a C^∞-vector field Y on M such that $X = Y$ on A.

Exercises

1. If f is a C^∞ function on a closed regular submanifold N of a manifold M, then show that f is the restriction of a C^∞ function on M.
2. Show that if N is a closed regular submanifold of M, then a C^∞-vector field X on N can be extended to a C^∞-vector field on M. [*Hint:* Take a covering of N by preferred coordinate neighborhoods of M and use a partition of unity subordinate to this covering and to the open set $M - N$; X can be extended easily within a preferred coordinate neighborhood.]

3. Show that on a Riemannian manifold every point p lies in an open set U_p over which we may define a C^∞ field of frames which is orthonormal at each point.
4. Let M be a manifold of dimension k and $F: M \to \boldsymbol{R}^n$ a C^∞ imbedding of M in $\boldsymbol{R}^n$. Further, let $G(n, k)$ be the Grassman manifold of k-planes through the origin of $\boldsymbol{R}^n$. Show that the map $H: M \to G(n, k)$, obtained by mapping p to the k-plane through the origin parallel to $F_*(T_p(M))$, is C^∞. [This generalizes the Gauss mapping for surfaces in $\boldsymbol{R}^3$.]
5. Show that if F_0 and F_1 are disjoint closed subsets of a C^∞ manifold M, then there exists a C^∞ function f on M that is 0 on F_0 and $+1$ on F_1 (compare Theorem III.3.4).
6. Prove Theorem 4.8 under the weaker assumption that A is a closed subset of M.
7. Extend Theorem 4.8 to a mapping into $\boldsymbol{R}^N$, arbitrary N, or into another manifold.
8. Prove that a continuous vector field can be approximated by a C^∞ vector field.
9. Prove that on every C^∞ manifold M it is possible to define a proper C^∞ function $f: M \to \boldsymbol{R}$, i.e., such that the inverse image of a compact set is compact.

5 Tensor Fields

Tensors on a Vector Space

It is our purpose in this section to define and study some properties of tensor fields on a manifold, especially covariant tensor fields. As in the case of covectors and bilinear forms, which are examples of such tensors, we begin with a vector space V over a field, in fact over $\boldsymbol{R}$.

(5.1) Definition A *tensor* Φ *on* V is by definition a multilinear map

$$\Phi: \underbrace{V \times \cdots \times V}_{r} \times \underbrace{V^* \times \cdots \times V^*}_{s} \to \boldsymbol{R}$$

V^* denoting the dual space to V, r its *covariant order*, and s its *contravariant order*. (Assume $r > 0$ or $s > 0$.)

Thus Φ assigns to each r-tuple of elements of V and s-tuple of elements of V^* a real number and if for each k, $1 \le k \le r + s$, we hold every variable except the kth fixed, then Φ satisfies the linearity condition

$$\Phi(\mathbf{v}_1, \ldots, \alpha\mathbf{v}_k + \alpha'\mathbf{v}'_k, \ldots) = \alpha\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_k, \ldots) + \alpha'\Phi(\mathbf{v}_1, \ldots, \mathbf{v}'_k, \ldots)$$

for all $\alpha, \alpha' \in \boldsymbol{R}$, and $\mathbf{v}_k, \mathbf{v}'_k \in V$ (or V^*, respectively). (This equation defines precisely the meaning of multilinearity.) As examples we have: (i) for $r = 1$,

$s = 0$, any $\varphi \in V^*$, (ii) for $r = 2$, $s = 0$, any bilinear form Φ on V, and finally (iii) the natural pairing of V and V^*, that is, $(\mathbf{v}, \varphi) \to \langle \varphi, \mathbf{v} \rangle$ for the case $r = 1$, $s = 1$. We have also noted that V and $(V^*)^*$ are naturally isomorphic and thus may be identified so then each $\mathbf{v} \in V$ may be considered as a linear map of V^* to $\boldsymbol{R}$, that is, as a tensor with $r = 0$ and $s = 1$.

For a fixed (r, s) we let $\mathscr{T}^r_s(V)$ be the collection of *all* tensors on V of covariant order r and contravariant order s. We know that as functions from $V \times \cdots \times V \times V^* \times \cdots \times V^*$ to $\boldsymbol{R}$ they may be added and multiplied by scalars (elements of $\boldsymbol{R}$). (Indeed linear combinations of functions from any set to $\boldsymbol{R}$ are defined and are again functions from that set to $\boldsymbol{R}$, a circumstance that we have used on several occasions.) With this addition and scalar multiplication $\mathscr{T}^r_s(V)$ is a vector space, so that if $\Phi_1, \Phi_2 \in \mathscr{T}^r_s(V)$ and $\alpha_1, \alpha_2 \in \boldsymbol{R}$, then $\alpha_1 \Phi_1 + \alpha_2 \Phi_2$, defined in the way alluded to above, that is, by

$$(\alpha_1 \Phi_1 + \alpha_2 \Phi_2)(\mathbf{v}_1, \mathbf{v}_2, \ldots) = \alpha_1 \Phi_1(\mathbf{v}_1, \mathbf{v}_2, \ldots) + \alpha_2 \Phi_2(\mathbf{v}_1, \mathbf{v}_2, \ldots),$$

is multilinear, and therefore is in $\mathscr{T}^r_s(V)$. Thus $\mathscr{T}^r_s(V)$ has a natural vector space structure. In this connection we have the following theorem:

(5.2) Theorem *With the natural definitions of addition and multiplication by elements of $\boldsymbol{R}$ the set $\mathscr{T}^r_s(V)$ of all tensors of order (r, s) on V forms a vector space of dimension n^{r+s}.*

Proof We consider the case $s = 0$ only, that is, covariant tensors of fixed order r, and we let $\mathscr{T}^r(V)$, rather than $\mathscr{T}^r_0(V)$, denote the collection of all such tensors. If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of V, then $\Phi \in \mathscr{T}^r(V)$ is completely determined by its n^r values on the basis vectors. Indeed by multilinearity if we write $\mathbf{v}_i = \sum \alpha_i^j \mathbf{e}_j$, $i = 1, \ldots, r$, then the value of Φ is given by the formula

$$\text{(5.2}'\text{)} \qquad \Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \sum_{j_1, \ldots, j_r} \alpha_1^{j_1} \alpha_2^{j_2} \cdots \alpha_r^{j_r} \Phi(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r}),$$

the sum being over all $1 \leq j_1, \ldots, j_r \leq n$. The n^r numbers $\{\Phi(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r})\}$ are called the *components* of Φ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. We shall justify the terminology by showing that there is in fact a basis of $\mathscr{T}^r(V)$, determined by $\mathbf{e}_1, \ldots, \mathbf{e}_n$, with respect to which these are components of Φ. It is defined as follows: Let $\Omega^{j_1 \cdots j_r}$ be that element of $\mathscr{T}^r(V)$ whose values on the basis vectors are given by

$$\Omega^{j_1 \cdots j_r}(\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_r}) = \begin{cases} +1 & \text{if} \quad j_i = k_i \quad \text{for} \quad i = 1, \ldots, r, \\ \ \ 0 & \text{if} \quad j_i \neq k_i \quad \text{for some } i, \end{cases}$$

and whose values on an arbitrary r-tuple $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$ is defined by (5.2′); whence

$$\Omega^{j_1 \cdots j_r}(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \alpha_1^{j_1} \alpha_2^{j_2} \cdots \alpha_r^{j_r}.$$

This does define a tensor: multilinearity is a consequence of this formula, which is linear in the components of each $\mathbf{v}_i$. It is immediate that the n^r tensors so chosen are linearly independent: If

$$\sum_{j_1, \ldots, j_r} \gamma_{j_1} \cdots \gamma_{j_r} \Omega^{j_1 \cdots j_r} = 0,$$

then it follows that

$$\sum_{j_1, \ldots, j_r} \gamma_{j_1 \cdots j_r} \Omega^{j_1 \cdots j_r}(\mathbf{v}_1, \ldots, \mathbf{v}_r) = 0$$

for any choice of the variables $\mathbf{v}_1, \ldots, \mathbf{v}_r$. But from the definition of the $\Omega^{j_1 \cdots j_r}$ we see, by substituting in turn each combination $\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_r}$ of basis elements as variables, that every coefficient $\gamma_{k_1 \cdots k_r} = 0$.

However, we also find that every Φ is a linear combination of these tensors. Let $\varphi_{j_1 \cdots j_r} = \Phi(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r})$ and consider the element $\sum \varphi_{j_1 \cdots j_r} \Omega^{j_1 \cdots j_r}$ of $\mathscr{T}^r(V)$. Applying again the definition of $\Omega^{j_1 \cdots j_r}$, we see that this tensor and Φ take the same values on every set of basis elements, hence must be equal. This completes the proof for $\mathscr{T}^r(V)$. ∎

We remark that an easy extension of the argument using both $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and its dual basis $\omega^1, \ldots, \omega^n$ of V^* gives the general case $\mathscr{T}^r_s(V)$. Since we use covariant tensors in most of what follows, we will leave the more general treatment to the exercises and to the imagination of the reader.

Tensor Fields

It is easy to extend these ideas to manifolds following the pattern we established earlier.

(5.3) Definition A C^∞-*covariant tensor field of order* r on a C^∞ manifold M is a function Φ which assigns to each $p \in M$ an element Φ_p of $\mathscr{T}^r(T_p(M))$ and which has the additional property that given any $X_1, \ldots, X_r$, C^∞-vector fields on an open subset U of M, then $\Phi(X_1, \ldots, X_r)$ is a C^∞ function on U, defined by $\Phi(X_1, \ldots, X_r)(p) = \Phi_p(X_{1p}, \ldots, X_{rp})$. We denote by $\mathscr{T}^r(M)$ the set of all C^∞-covariant tensor fields of order r on M.

We have already considered in some detail the case of covector fields, $r = 1$, and that of fields of bilinear forms, $r = 2$. Just as in these cases, it is an immediate consequence of the definition that a covariant tensor field of order r is not only $\boldsymbol{R}$-linear but also $C^\infty(M)$-linear in each variable. For example, if $f \in C^\infty(M)$,

$$\Phi(X_1, \ldots, fX_i, \ldots, X_r) = f\Phi(X_1, \ldots, X_i, \ldots, X_r).$$

This is true because it holds at each point p by the $\boldsymbol{R}$-linearity of Φ_p; the two sides of the equation are equal since equality holds for each $p \in M$. In the same way, if $f \in C^\infty(U)$, U open in M, the equation holds for Φ_U, the restriction of Φ to U. (Compare Remark 1.3 and Exercise 2.2.)

In precisely the same fashion as Section 2 we see that if U, φ is a coordinate neighborhood and $E_1, \ldots, E_n$ are the coordinate frames, then $\Phi \in \mathscr{T}^r(M)$ has *components* $\Phi(E_{j_1}, \ldots, E_{j_r})$, C^∞ functions on U whose values at each $p \in U$ are the components of Φ_p relative to the basis of $T_p(M)$ determined by $E_1, \ldots, E_n$. Once more, just as before, the differentiability of Φ is implied by the differentiability of all the components as functions on the coordinate neighborhoods of some covering of M. Finally, it is easy to see that $\mathscr{T}^r(M)$ is a vector space over $\boldsymbol{R}$ [in fact it is a $C^\infty(M)$ module] since linear combinations of covariant tensors of order r (even with C^∞ functions as coefficients) are again covariant tensor fields.

Mappings and Covariant Tensors

A further basic fact which carries over to arbitrary $r > 0$ from covectors and forms is that any linear map of vector spaces $F_*: V \to W$ induces a linear map $F^*: \mathscr{T}^r(W) \to \mathscr{T}^r(V)$ by the formula

$$F^*\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \Phi(F_*(\mathbf{v}_1), \ldots, F_*(\mathbf{v}_r)).$$

In exact analogy with the case $r = 2$, we find that a C^∞-map $F: M \to N$ induces a mapping $F^*: \mathscr{T}^r(N) \to \mathscr{T}^r(M)$, defined for Φ on N by

$$(F^*\Phi)_p(X_{1p}, \ldots, X_{rp}) = \Phi_{F(p)}(F_*(X_{1p}), \ldots, F_*(X_{rp})).$$

As we have seen, this is a special feature of *covariant* tensor fields; its analog does not hold for *contravariant* fields even for $\mathscr{T}_1(M) = \mathfrak{X}(M)$ (vector fields); see Definition IV.2.6. Not only does F^* map $\mathscr{T}^r(N)$ to $\mathscr{T}^r(M)$ but it maps it linearly; this is an immediate consequence of the definitions (compare Exercise V.2.6).

(5.4) Definition We shall say that $\Phi \in \mathscr{T}^r(V)$, V a vector space, is *symmetric* if for each $1 \le i, j \le r$, we have

$$\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_r) = \Phi(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_r).$$

Similarly, if interchanging the ith and jth variables, $1 \le i, j \le r$, changes the sign,

$$\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_r) = -\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_r),$$

then we say Φ is *skew* or *antisymmetric* or *alternating*; alternating covariant tensors are often called *exterior forms*. A tensor field is symmetric (respectively, alternating) if it has this property at each point.

The following generalization of Theorems 1.6 and 2.3 summarizes these remarks:

(5.5) Theorem *Let $F: M \to N$ be a C^∞ map of C^∞ manifolds. Then each C^∞-covariant tensor field Φ on N determines a C^∞-covariant tensor field $F^*\Phi$ on M by the formula*

(5.5′) $$(F^*\Phi)_p(X_{1p}, \ldots, X_{rp}) = \Phi_{F(p)}(F_*(X_{1p}), \ldots, F_*(X_{rp})).$$

The map $F^*: \mathscr{T}^r(N) \to \mathscr{T}^r(M)$ so defined is linear and takes symmetric (alternating) tensors to symmetric (alternating) tensors.

We leave the proof as an exercise. Note that (5.5′) is the same as formula (2.2). [Warning: under composition we have $(F \circ G)^* = G^* \circ F^*$.]

It is also clear how to extend to the case of arbitrary order r the formula (2.4) for components of $F^*\Phi$ in terms of those of Φ and the Jacobian of F in local coordinates. The same method can also be used to derive formulas for change of components relative to a change of local coordinates (for $r = 1$ see Corollary 1.7). Basically, these formulas are all consequences of the multilinearity *at each point* of M.

The Symmetrizing and Alternating Transformations

In order to pursue some of these questions somewhat further, we return to the case of a covariant tensor on a vector space V. First note that if Φ_1 and $\Phi_2 \in \mathscr{T}^r(V)$ are symmetric (respectively, alternating) covariant tensors of order r on V, then a linear combination $\alpha\Phi_1 + \beta\Phi_2$, $\alpha, \beta \in \boldsymbol{R}$, is also symmetric (respectively, alternating). Thus the symmetric tensors in $\mathscr{T}^r(V)$ form a subspace which we denote by $\Sigma^r(V)$ and the alternating tensors (exterior forms) also form a subspace $\bigwedge^r(V)$. These subspaces have only the 0-tensor in common.

Next let σ denote a permutation of $(1, \ldots, r)$ with $(1, \ldots, r) \to (\sigma(1), \ldots, \sigma(r))$. We know that any such permutation is a product of permutations interchanging just two elements (transpositions). Although this representation is not unique the parity (evenness or oddness of the number of factors) is. We let sgn $\sigma = +1$ if σ is representable as the product of an even number of transpositions and sgn $\sigma = -1$ otherwise. Then, $\sigma \to \operatorname{sgn} \sigma$ is a well-defined map from the group of permutations of r letters $\mathfrak{S}_r$ to the multiplicative group of two elements ± 1. It is even a homomorphism as we check at once from the definition. The only statement not obvious is the one concerning the independence of the parity of the particular decomposition of σ into a product of transpositions (for a proof see Zassenhaus [1]).

In the light of these facts we see that our original definitions may be restated in the following equivalent form: $\Phi \in \mathscr{T}^r(V)$ is *symmetric* if

$\Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}) = \Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ for every $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and permutation σ, and is *alternating* if $\Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}) = \operatorname{sgn} \sigma \Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ for every $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and permutation σ.

(5.6) Definition We define two linear transformations on the vector space $\mathscr{T}^r(V)$,

$$\text{symmetrizing mapping} \qquad \mathscr{S}: \mathscr{T}^r(V) \to \mathscr{T}^r(V),$$

$$\text{alternating mapping} \qquad \mathscr{A}: \mathscr{T}^r(V) \to \mathscr{T}^r(V),$$

by the formulas:

$$(\mathscr{S}\Phi)(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \frac{1}{r!} \sum_\sigma \Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)})$$

and

$$(\mathscr{A}\Phi)(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \frac{1}{r!} \sum_\sigma \operatorname{sgn} \sigma \Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}),$$

the summation being over all $\sigma \in \mathfrak{S}_r$, the group of all permutations of r letters.

It is immediate that these maps are linear transformations on $\mathscr{T}^r(V)$, in fact $\Phi \to \Phi^\sigma$, defined by

$$\Phi^\sigma(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}),$$

is such a linear transformation; and any linear combination of linear transformations of a vector space is again a linear transformation. We have the following properties.

(5.7) *Properties of $\mathscr{A}$ and $\mathscr{S}$:*

(i) *$\mathscr{A}$ and $\mathscr{S}$ are projections*, that is, *$\mathscr{A}^2 = \mathscr{A}$ and $\mathscr{S}^2 = \mathscr{S}$;*
(ii) *$\mathscr{A}(\mathscr{T}^r(V)) = \bigwedge^r(V)$ and $\mathscr{S}(\mathscr{T}^r(V)) = \Sigma^r(V)$;*
(iii) *Φ is alternating if and only if $\mathscr{A}\Phi = \Phi$;*
Φ is symmetric if and only if $\mathscr{S}\Phi = \Phi$;
(iv) *if $F_*: V \to W$ is a linear map, then $\mathscr{A}$ and $\mathscr{S}$ commute with $F^*: \mathscr{T}^r(W) \to \mathscr{T}^r(V)$.*

All of these statements are easy consequences of the definitions. We shall check them only for $\mathscr{A}$, the verification for $\mathscr{S}$ being similar. They are also interrelated so we will not take them in order. First note that if Φ is alternating, then the definition and $(\operatorname{sgn} \sigma)^2 = 1$ imply

$$\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \operatorname{sgn} \sigma \Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}).$$

Since there are $r!$ elements of $\mathfrak{S}_r$, summing both sides over all $\sigma \in \mathfrak{S}_r$ gives $\Phi = \mathscr{A}\Phi$. On the other hand if we apply a permutation τ to the variables of $\mathscr{A}\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ for an arbitrary $\Phi \in \mathscr{T}^r(V)$, we obtain

$$\mathscr{A}\Phi(\mathbf{v}_{\tau(1)}, \ldots, \mathbf{v}_{\tau(r)}) = \frac{1}{r!} \sum_{\sigma} \operatorname{sgn} \sigma \Phi(\mathbf{v}_{\sigma\tau(1)}, \ldots, \mathbf{v}_{\sigma\tau(r)}).$$

Now sgn is a homomorphism and $\operatorname{sgn} \tau^2 = 1$ so that $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma\tau \operatorname{sgn} \tau$. From this equation and since $\sigma\tau$ runs through $\mathfrak{S}_r$ as σ does, we see that the right side is

$$\frac{1}{r!} \operatorname{sgn} \tau \sum_{\sigma} \operatorname{sgn} \sigma\tau \Phi(\mathbf{v}_{\sigma\tau(1)}, \ldots, \mathbf{v}_{\sigma\tau(r)}) = \operatorname{sgn} \tau \mathscr{A}\Phi(\mathbf{v}_1, \ldots, \mathbf{v}_r),$$

and $\mathscr{A}\Phi$ is alternating. This shows that $\mathscr{A}(\mathscr{T}^r(V)) \subset \bigwedge^r(V)$. If Φ is alternating, every term in the summation defining $\mathscr{A}\Phi$ is equal, so $\mathscr{A}\Phi = \Phi$. Thus $\mathscr{A}$ is the identity on $\bigwedge^r(V)$ and $\mathscr{A}(\mathscr{T}^r(V)) \supset \bigwedge^r(V)$. From these facts (i)–(iii) all follow for $\mathscr{A}$. Statement (iv) is immediate from the definition of F^*, for we have

$$F^*\Phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}) = \Phi(F_*(\mathbf{v}_{\sigma(1)}), \ldots, F_*(\mathbf{v}_{\sigma(r)})).$$

Multiplying both sides by sgn σ and summing over all σ gives—if we use the linearity of F^*—$\mathscr{A}(F^*\Phi)(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ on the left and $F^*(\mathscr{A}\Phi)(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ on the right.

Both of these maps $\mathscr{A}$ and $\mathscr{S}$ can be immediately extended to mappings of tensor fields on manifolds—with the same properties—by merely applying them at each point and then verifying that both sides of each relation (i)–(iv) give C^∞ functions which agree pointwise on every r-tuple of C^∞-vector fields. We summarize (without proof):

(5.8) Theorem *The maps $\mathscr{A}$ and $\mathscr{S}$ are defined on $\mathscr{T}^r(M)$ (M a C^∞ manifold and $\mathscr{T}^r(M)$ the C^∞-covariant tensor fields of order r) and they satisfy properties* (5.7), (i)–(iv), *there. In the case of* (iv), *$F^*: \mathscr{T}^r(N) \to \mathscr{T}^r(M)$ is the linear map induced by a C^∞ mapping $F: M \to N$.*

Exercises

1. Assume that $\dim V > 1$ and that when $r = 2$ we have $\mathscr{T}^r(V) = \bigwedge^r(V) \oplus \sum^r(V)$ but that this is in general false if $r > 2$.
2. Show that $\bigwedge^r(V)$ contains only the tensor 0 when $r > \dim V$.
3. Let $\mathscr{E}(V)$ denote the space of all linear transformations on V. Show that it is a vector space over $\boldsymbol{R}$ of dimension equal to $(\dim V)^2$ and that it is naturally isomorphic to the space of all bilinear maps of $V \times V^*$ to $\boldsymbol{R}$. [*Hint:* To a linear transformation $A: V \to V$ we associate the bilinear map $(\mathbf{v}, \varphi) \to \langle A\mathbf{v}, \varphi \rangle$ on $V \times V^*$ to $\boldsymbol{R}$.]

4. Give a definition of a C^∞ field of *linear transformations* on M and check that its property of being C^∞ can be defined in terms of local coordinates or in terms of $\mathfrak{X}(M)$.
5. If V is a vector space with an inner product, then there is a natural isomorphism of V to V^* (compare (iii) of Section 2). Show that this determines an isomorphism of $\mathscr{T}^r_s(V)$ and $\mathscr{T}^{r+s}(V)$ and extend to tensor fields on C^∞ manifolds.
6. If Φ is a C^∞-covariant tensor field of order r on a C^∞ manifold M, show that $\Phi(X_1, \ldots, X_r)$ is a $C^\infty(M)$-linear function from $\mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M)$ to $C^\infty(M)$. Conversely, show that each such function determines an element of $\mathscr{T}^r(M)$ as we have defined it.
7. Let $T\colon V \times \cdots \times V \to V$ be an $\boldsymbol{R}$-linear function of r vectors with values in V, that is, $T(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ is in V and linear in each variable. Define components in this case and extend this object to a field on a manifold in the manner of Exercises 3 and 4.
8. As in the case of vector fields on manifolds, a tensor field, say $\Phi \in \mathscr{T}^r(M)$, is a function assigning to each $p \in M$ a covariant tensor Φ_p on $T_p(M)$, that is, a function from M to the set $W = \bigcup_{p \in M} \mathscr{T}^r(T_p(M))$. Try to define the structure of a C^∞ manifold on W such that (1) the natural mapping $\pi\colon W \to M$ taking Φ_p to p for each $\Phi_p \in \mathscr{T}^r(T_p(M))$ is C^∞ and (2) covariant tensor fields on M are exactly the C^∞ mappings $\Phi\colon M \to W$ satisfying $\pi \circ \Phi =$ identity (on M).

6 Multiplication of Tensors

Except for a few of the exercises, we will continue to restrict our attention to *covariant* tensors in the remainder of this chapter and in the next. Thus V will denote a vector space and M a C^∞ manifold, as before. We have seen that both $\mathscr{T}^r(V)$ and $\mathscr{T}^r(M)$ are vector spaces over $\boldsymbol{R}$. In the case of tensor fields, $\mathscr{T}^r(M)$ has also the structure of a $C^\infty(M)$-module. We agree by definition that $\mathscr{T}^0(V) = \boldsymbol{R}$ and $\mathscr{T}^0(M) = C^\infty(M)$. Having made these conventions, recall that our viewpoint is to define tensors as functions to $\boldsymbol{R}$, a field, in the case of $\mathscr{T}^r(V)$ and functions to $C^\infty(M)$, an algebra, in the case of $\mathscr{T}^r(M)$. In either case it is appropriate to discuss products of such functions. Just as functions from a set to an algebra can be multiplied in a natural way—using the algebra product of their values—to give new functions of the same type, so can we hope to multiply tensors. As usual we begin with the vector space case.

Multiplication of Tensors on a Vector Space

Suppose V is a vector space and $\varphi \in \mathscr{T}^r(V)$, $\psi \in \mathscr{T}^s(V)$ are tensors. Their product is easily seen to be linear in each of its $r + s$ variables so we make the following definition:

(6.1) Definition The *product* of φ and ψ, denoted $\varphi \otimes \psi$ is a tensor of order $r + s$ defined by

$$\varphi \otimes \psi(\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_{r+s}) = \varphi(\mathbf{v}_1, \ldots, \mathbf{v}_r)\psi(\mathbf{v}_{r+1}, \ldots, \mathbf{v}_{r+s}).$$

The right-hand side is the product of the values of φ and ψ. The product defines a mapping $(\varphi, \psi) \to \varphi \otimes \psi$ of $\mathscr{T}^r(V) \times \mathscr{T}^s(V) \to \mathscr{T}^{r+s}(V)$.

(6.2) Theorem *The product $\mathscr{T}^r(V) \times \mathscr{T}^s(V) \to \mathscr{T}^{r+s}(V)$ just defined is bilinear and associative. If $\omega^1, \ldots, \omega^n$ is a basis of $V^* = \mathscr{T}^1(V)$, then $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}\}$ over all $1 \le i_1, \ldots, i_r \le n$ is a basis of $\mathscr{T}^r(V)$. Finally, if $F_*: W \to V$ is linear, then $F^*(\varphi \otimes \psi) = (F^*\varphi) \otimes (F^*\psi)$.*

Proof Each statement is proved by straightforward computation. To say that $\otimes$ is *bilinear* means that if α, β, are numbers $\varphi_1, \varphi_2 \in \mathscr{T}^r(V)$ and $\psi \in \mathscr{T}^s(V)$, then $(\alpha\varphi_1 + \beta\psi_2) \otimes \psi = \alpha(\varphi_1 \otimes \psi) + \beta(\varphi_2 \otimes \psi)$. Similarly for the second variable. This is checked by evaluating each side on $r + s$ vectors of V; in fact basis vectors suffice because of linearity. Associativity, $(\varphi \otimes \psi) \otimes \theta = \varphi \otimes (\psi \otimes \theta)$, is similarly verified—the products on both sides being defined in the natural way. This allows us to drop the parentheses. To see that $\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}$ form a basis it is sufficient to note that if $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the basis of V dual to $\omega^1, \ldots, \omega^n$, then the tensor $\Omega^{i_1 \cdots i_r}$ previously defined is exactly $\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}$. This follows from the two definitions:

$$\Omega^{i_1 \cdots i_r}(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r}) = \begin{cases} 0 & \text{if} \quad (i_1, \ldots, i_r) \neq (j_1 \cdots j_r), \\ 1 & \text{if} \quad (i_1, \ldots, i_r) = (j_1 \cdots j_r), \end{cases}$$

and

$$\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r}) = \omega^{i_2}(\mathbf{e}_{j_1})\omega^{i_2}(\mathbf{e}_{j_2}) \cdots \omega^{i_2}(\mathbf{e}_{j_r}) = \delta^{i_1}_{j_1}\,\delta^{i_2}_{j_2} \cdots \delta^{i_r}_{j_r},$$

which show that both tensors have the same values on any (ordered) set of r basis vectors and are thus equal.

Finally, given $F_*: W \to V$, if $\mathbf{w}_1, \ldots, \mathbf{w}_{r+s} \in W$, then

$$\begin{aligned}(F^*(\varphi \otimes \psi))(\mathbf{w}_1, \ldots, \mathbf{w}_{r+s}) &= \varphi \otimes \psi(F_*(\mathbf{w}_1), \ldots, F_*(\mathbf{w}_{r+s})) \\ &= \varphi(F_*(\mathbf{w}_1), \ldots, F_*(\mathbf{w}_r))\psi(F_*(\mathbf{w}_{r+1}), \ldots, F_*(\mathbf{w}_{r+s})) \\ &= (F^*\varphi) \otimes (F^*\psi)(\mathbf{w}_1, \ldots, \mathbf{w}_{r+s}),\end{aligned}$$

which proves $F^*(\varphi \otimes \psi) = (F^*\varphi) \otimes (F^*\psi)$ and completes the proof. ∎

A more sophisticated but no more general way to state this theorem is derived from the following observations. First we take the direct sum over $\boldsymbol{R}$ of all of the tensor spaces, beginning with $\mathscr{T}^0(V) = \boldsymbol{R}$. We denote it by $\mathscr{T}(V)$,

$$\mathscr{T}(V) = \mathscr{T}^0(V) \oplus \mathscr{T}^1(V) \oplus \cdots \oplus \mathscr{T}^r(V) \oplus \cdots.$$

We identify each $\mathscr{T}^r(V)$ with its (natural) isomorphic image in $\mathscr{T}(V)$. An element φ of $\mathscr{T}(V)$ is said to be of *order r* if it is in $\mathscr{T}^r(V)$, and every element $\tilde{\varphi}$ of $\mathscr{T}(V)$ is the sum of a finite number of such φ, which we call its *components*. Thus $\tilde{\varphi} \in \mathscr{T}(V)$ may be written uniquely $\tilde{\varphi} = \varphi_1^{i_1} + \cdots + \varphi_n^{i_n}$, where $\varphi^{i_j} \in \mathscr{T}^{i_j}(V)$ and $i_1 < i_2 < \cdots < i_r$. If $\tilde{\varphi}, \tilde{\psi} \in \mathscr{T}(V)$, then they may be added componentwise, that is, by adding in $\mathscr{T}^r(V)$ any terms in $\mathscr{T}^r(V)$. They may be multiplied by using $\otimes$, extending it to be distributive on all of $\mathscr{T}(V)$. This makes $\mathscr{T}(V)$ into an associative algebra over $\boldsymbol{R}$ called the *tensor algebra*. It contains $\boldsymbol{R} = \mathscr{T}^0(V)$, has 1 as its unit, and is infinite-dimensional. The contents of Theorem 6.2 (even a little more) can be written:

(6.2′) $\mathscr{T}(V) = \sum_{r=0}^{\infty} \mathscr{T}^r(V)$ *(direct) is an associative algebra (with unit) over* $R = \mathscr{T}^0(V)$. *It is generated by* $\mathscr{T}^0(V)$ *and* $\mathscr{T}^1(V) = V^*$, *the dual space to* V. *Any linear mapping* $F_*\colon W \to V$ *of vector spaces induces a homomorphism* $F^*\colon \mathscr{T}(V) \to \mathscr{T}(W)$ *which is (i) the identity on* $\boldsymbol{R}$ *and (ii) the dual mapping* $F^*\colon V^* \to W^*$ *on* $\mathscr{T}^1(V)$. *Together (i) and (ii) determine* F^* *uniquely on all of* $\mathscr{T}(V)$.

Multiplication of Tensor Fields

Now we turn briefly to the case of tensor fields on a manifold M. If $\varphi \in \mathscr{T}^r(M)$ and $\psi \in \mathscr{T}^s(M)$, then we may define $\varphi \otimes \psi$ on M by defining it at each point using the definition for tensors on a vector space, that is, $(\varphi \otimes \psi)_p$ is defined to be the tensor $\varphi_p \otimes \psi_p$ of order $r + s$ on the vector space $T_p(M)$. Since this defines a covariant tensor of order $r + s$ on the tangent space at each point of M, it will define a tensor field—if it is C^∞. Now in local coordinates the components of $\varphi \otimes \psi$, according to the definition just given, are the functions of the coordinate frame vectors

$$\varphi \otimes \psi(E_{i_1}, \ldots, E_{i_{r+s}}) = \varphi(E_{i_1}, \ldots, E_{i_r})\psi(E_{i_{r+1}}, \ldots, E_{i_{r+s}})$$

over the coordinate neighborhood. The right-hand side is the product of two C^∞ functions, components in local coordinates of φ and ψ, and thus the left side is C^∞ as hoped. We have an appropriate version of Theorem 6.2 for this case.

(6.3) Theorem *The mapping* $\mathscr{T}^r(M) \times \mathscr{T}^s(M) \to \mathscr{T}^{r+s}(M)$ *just defined is bilinear and associative. If* $\omega^1, \ldots, \omega^n$ *is a basis of* $\mathscr{T}^1(M)$, *then every element of* $\mathscr{T}^r(M)$ *is a linear combination with* C^∞ *coefficients of* $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r} \mid 1 \le i_1, \ldots, i_r \le n\}$. *If* $F\colon N \to M$ *is a* C^∞ *mapping,* $\varphi \in \mathscr{T}^r(M)$ *and* $\psi \in \mathscr{T}^s(M)$, *then* $F^*(\varphi \otimes \psi) = (F^*\varphi) \otimes (F^*\psi)$, *tensor fields on* N.

Proof Since two tensor fields are equal if and only if they are equal at each point, it is only necessary to see that these equations hold at each point, which follows at once from the definitions and the preceding Theorem 6.2. ∎

In general we do not have a globally defined basis of $\mathscr{T}^1(M)$, that is, covector fields $\omega^1, \ldots, \omega^n$, which are a basis at each point. However, we do in $\boldsymbol{R}^n$, from which the following corollary is obtained by applying the theorem to a coordinate neighborhood U, θ of M. Let $E_1, \ldots, E_n$ denote the coordinate frames and $\omega^1, \ldots, \omega^n$ their duals, that is, $E_i = \theta_*^{-1}(\partial/\partial x^i)$ and $\omega^j = \theta^*(dx^j)$.

(6.4) Corollary *Each $\varphi \in \mathscr{T}^r(U)$, including the restriction to U of any covariant tensor field on M, has a unique expression of the form*

$$\varphi = \sum_{i_1} \cdots \sum_{i_r} a_{i_1 \cdots i_r} \omega^{i_1} \otimes \cdots \otimes \omega^{i_r},$$

where at each point of U, $a_{i_1 \cdots i_r} = \varphi(E_{i_1}, \ldots, E_{i_r})$ are the components of φ in the basis $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}\}$ and are C^∞ functions on U.

Exterior Multiplication of Alternating Tensors

For each $r > 0$ we have defined the subspace $\bigwedge^r(V) \subset \mathscr{T}^r(V)$ consisting of alternating covariant tensors of order r; it is the image of $\mathscr{T}^r(V)$ under the linear mapping $\mathscr{A}$, the alternating mapping. We define $\bigwedge^0(V)$ to be $\boldsymbol{R}$, the field. Then $\bigwedge^0(V) = \mathscr{T}^0(V) = \boldsymbol{R}$ and $\bigwedge^1(V) = \mathscr{T}^1(V) = V^*$, but $\bigwedge^r(V)$ is properly contained in $\mathscr{T}^r(V)$ for $r > 1$ (Exercise 5.2). We see, therefore, that the direct sum $\bigwedge(V)$ of all the spaces $\bigwedge^r(V)$ is contained in $\mathscr{T}(V)$ as a subspace:

$$\begin{aligned}\textstyle\bigwedge(V) &= \textstyle\bigwedge^0(V) \oplus \bigwedge^1(V) \oplus \bigwedge^2(V) \oplus \cdots \\ &\subset \mathscr{T}^0(V) \oplus \mathscr{T}^1(V) \oplus \mathscr{T}^2(V) \oplus \cdots = \mathscr{T}(V).\end{aligned}$$

Although $\bigwedge(V)$ is a *subspace* of $\mathscr{T}(V)$, it is not a *subalgebra*. For even if $\varphi \in \bigwedge^r(V)$ and $\psi \in \bigwedge^s(V)$, it may be shown by example (Exercise 1) that $\varphi \otimes \psi$ may very well fail to be an element of $\bigwedge^{r+s}(V)$; thus the tensor product of alternating tensors on V is not, in general, an alternating tensor on V. We know, however, that each tensor determines an alternating tensor, its image under $\mathscr{A}$. This fact enables us to define another multiplication for alternating tensors that is extraordinarily useful.

(6.5) Definition The mapping from $\bigwedge^r(V) \times \bigwedge^s(V) \to \bigwedge^{r+s}(V)$, defined by

$$(\varphi, \psi) \to \frac{(r+s)!}{r!\,s!} \mathscr{A}(\varphi \otimes \psi),$$

is called the *exterior product* (or *wedge product*) of φ and ψ and is denoted $\varphi \wedge \psi$.

(6.6) Lemma *The exterior product is bilinear and associative.*

Proof Bilinearity is a consequence of the fact that the product is defined by composing the tensor product, a bilinear mapping from $\bigwedge^r(V) \times \bigwedge^s(V)$ to $\mathscr{T}^{r+s}(V)$ with a linear mapping $((r+s)!/r!\,s!)\mathscr{A}$.

To show that the product is associative we first prove a property of the alternating mapping $\mathscr{A}$. Suppose $\varphi \in \mathscr{T}^r(V)$, $\psi \in \mathscr{T}^s(V)$, and $\theta \in \mathscr{T}^t(V)$. Then we show that

$$\mathscr{A}(\varphi \otimes \psi \otimes \theta) = \mathscr{A}(\mathscr{A}(\varphi \otimes \psi) \otimes \theta) = \mathscr{A}(\varphi \otimes \mathscr{A}(\psi \otimes \theta)).$$

For this purpose let $\mathfrak{S} = \mathfrak{S}_{r+s+t}$ denote the permutations of $(1, 2, \ldots, r+s+t)$ and $\mathfrak{S}'$ the subgroup which leaves the last t integers fixed. Denote by $\sigma \to \sigma'$ the isomorphism of the permutation group $\mathfrak{S}_{r+s}$ of $(1, 2, \ldots, r+s)$ onto $\mathfrak{S}'$. We then have

$$\begin{aligned}
&\mathscr{A}(\varphi \otimes \psi) \otimes \theta(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s+t})\\
&\quad = \frac{1}{(r+s)!}\left(\sum_{\sigma \in \mathfrak{S}_{rs}} \operatorname{sgn} \sigma\varphi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)})\psi(\mathbf{v}_{\sigma(r+1)}, \ldots, \mathbf{v}_{\sigma(r+s)})\right)\\
&\qquad \cdot \theta(\mathbf{v}_{r+s+1}, \ldots, \mathbf{v}_{r+s+t})\\
&\quad = \frac{1}{(r+s)!}\sum_{\sigma' \in \mathfrak{S}'} \operatorname{sgn} \sigma'\varphi(\mathbf{v}_{\sigma(1)}, \ldots)\psi(\mathbf{v}_{\sigma(r+1)}, \ldots)\\
&\qquad \cdot \theta(\mathbf{v}_{\sigma(r+s+1)}, \ldots, \mathbf{v}_{\sigma(r+s+t)}).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathscr{A}(\mathscr{A}(\varphi \otimes \psi) \otimes \theta)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s+t})\\
&\quad = \frac{1}{(r+s+t)!}\,\frac{1}{(r+s)!}\\
&\qquad \cdot \sum_{\tau \in \mathfrak{S}} \sum_{\sigma' \in \mathfrak{S}'} \operatorname{sgn} \tau\sigma'\varphi(\mathbf{v}_{\tau\sigma'(1)}, \ldots)\psi(\mathbf{v}_{\tau\sigma'(r+1)}, \ldots)\theta(\mathbf{v}_{\tau\sigma'(r+s+1)}, \ldots).
\end{aligned}$$

Reversing the order of summation, i.e., summing over τ, yields

$$\begin{aligned}
&\frac{1}{(r+s)!}\sum_{\sigma' \in \mathfrak{S}} \mathscr{A}(\varphi \otimes \psi \otimes \theta)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s+t})\\
&\qquad = \mathscr{A}(\varphi \otimes \psi \otimes \theta)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s+t})
\end{aligned}$$

since the sum on the left contains $(r+s)!$ identical summands. Thus, we obtain

$$\mathscr{A}(\varphi \otimes \psi \otimes \theta) = \mathscr{A}(\mathscr{A}(\varphi \otimes \psi) \otimes \theta).$$

The second equality is proved in the same way. If φ, ψ, θ are in the subspaces $\bigwedge^r(V)$, $\bigwedge^s(V)$, $\bigwedge^t(V)$, respectively, then by definition

$$\varphi \wedge \psi = \frac{(r+s)!}{r!\,s!}\mathscr{A}(\varphi \otimes \psi)$$

and

$$(\varphi \wedge \psi) \wedge \theta = \frac{(r+s+t)!}{(r+s)!\,t!}\mathscr{A}\big((\varphi \wedge \psi) \otimes \theta\big).$$

From this and a similar expression in the other order of associating terms we obtain the associativity of the exterior product

$$(\varphi \wedge \psi) \wedge \theta = \varphi \wedge (\psi \wedge \theta). \quad \blacksquare$$

The following relation is an immediate consequence of the proof, which allows us to write exterior products without parentheses.

(6.7) Corollary *Let* $\varphi_i \in \bigwedge^{r_i}(V)$, $i = 1, \ldots, k$. *Then*

$$\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_k = \frac{(r_1 + r_2 + \cdots + r_k)!}{r_1!\,r_2!\cdots r_k!}\mathscr{A}(\varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_k).$$

Lemma 6.6 makes it possible for us to give $\bigwedge(V)$ the structure of an associative algebra over $\boldsymbol{R}$; we define the product $\bigwedge(V) \times \bigwedge(V) \to \bigwedge(V)$ simply by extending the exterior product to be bilinear, so that the distributive law holds. This is possible in only one way: Suppose that $\varphi, \psi \in \bigwedge(V)$. Then

$$\varphi = \varphi_1 + \cdots + \varphi_k, \quad \varphi_i \in \textstyle\bigwedge^{r_i}(V), \qquad \psi = \psi_1 + \cdots + \psi_l, \quad \psi_l \in \textstyle\bigwedge^{s_l}(V),$$

and we define

$$\varphi \wedge \psi = \sum_{i=1}^{k}\sum_{j=1}^{l} \varphi_i \wedge \psi_j .$$

(6.8) Corollary $\bigwedge(V) = \bigwedge^0(V) \oplus \bigwedge^1(V) \oplus \bigwedge^2(V) \oplus \cdots$ *with the exterior product as defined above is an (associative) algebra over* $\boldsymbol{R} = \bigwedge^0(V)$.

The algebra $\bigwedge(V)$ is called the *exterior algebra* or *Grassman algebra* over V. Unlike the tensor algebra $\mathscr{T}(V)$, of which it is a subspace (but not a subalgebra), it is finite-dimensional. To see this we determine a basis of $\bigwedge(V)$ as a vector space. For this we need the following lemma, which is important in its own right.

(6.9) Lemma *If $\varphi \in \bigwedge^r(V)$ and $\psi \in \bigwedge^s(V)$, then*

$$\varphi \wedge \psi = (-1)^{rs}\psi \wedge \varphi.$$

Proof This is equivalent to showing that

$$\mathscr{A}(\varphi \otimes \psi) = (-1)^{rs}\mathscr{A}(\psi \otimes \varphi).$$

To prove this equality we note that

$$\begin{aligned}
&\mathscr{A}(\varphi \otimes \psi)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s}) \\
&\quad = \frac{1}{(r+s)!}\sum_{\sigma} \operatorname{sgn} \sigma \varphi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)})\psi(\mathbf{v}_{\sigma(r+1)}, \ldots, \mathbf{v}_{\sigma(r+s)}) \\
&\quad = \frac{1}{(r+s)!}\sum_{\sigma} \operatorname{sgn} \sigma \psi(\mathbf{v}_{\sigma(r+1)}, \ldots, \mathbf{v}_{\sigma(r+s)})\varphi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(r)}).
\end{aligned}$$

If τ is the permutation taking $(1, \ldots, r, r+1, \ldots, r+s)$ to $(r+1, \ldots, r+s, 1, \ldots, r)$, then we may write, using $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma\tau \operatorname{sgn} \tau$,

$$\begin{aligned}
&\mathscr{A}(\varphi \otimes \psi)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s}) \\
&\quad = \frac{1}{(r+s)!}\sum_{\sigma} \operatorname{sgn} \sigma\tau \operatorname{sgn} \tau\psi(\mathbf{v}_{\sigma\tau(1)}, \ldots, \mathbf{v}_{\sigma\tau(s)})\varphi(\mathbf{v}_{\sigma\tau(s+1)}, \ldots, \mathbf{v}_{\sigma\tau(r+s)}) \\
&\quad = \operatorname{sgn} \tau \mathscr{A}(\psi \otimes \varphi)(\mathbf{v}_1, \ldots, \mathbf{v}_{r+s}).
\end{aligned}$$

Since it is easily checked that $\operatorname{sgn} \tau = (-1)^{rs}$, this gives the relation $\varphi \wedge \psi = (-1)^{rs}\psi \wedge \varphi$. ∎

(6.10) Theorem *If $r > n = \dim V$, then $\bigwedge^r(V) = \{0\}$. For $0 \leq r \leq n$, $\dim \bigwedge^r(V) = \binom{n}{r}$. Let $\omega^1, \ldots, \omega^n$ be a basis of $\bigwedge^1(V)$. Then the set*

$$\{\omega^{i_1} \wedge \cdots \wedge \omega^{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

is a basis of $\bigwedge^r(V)$ and $\dim \bigwedge(V) = 2^n$.

Proof Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be any basis of V. If φ is an alternating covariant tensor of order $r > \dim V$, then on any set of basis elements $\varphi(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_r}) = 0$. For some variable $\mathbf{e}_{i_k}$ is repeated and interchanging two equal variables both changes the sign of φ on the set and leaves it unchanged—the same argument one uses to show that a determinant with two equal rows is zero. Since all components of φ are zero, $\varphi = 0$ so $\bigwedge^r(V) = \{0\}$.

Suppose that $0 \leq r \leq n$ and that $\omega_1, \ldots, \omega_n$ is the basis of $V^* = \bigwedge^1(V)$ dual to $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Since $\mathscr{A}$ maps $\mathscr{T}^r(V)$ onto $\bigwedge^r(V)$, the image of the basis $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}\}$ of $\mathscr{T}^r(V)$ spans $\bigwedge^r(V)$. We have

$$r!\,\mathscr{A}(\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}) = \omega^{i_1} \wedge \cdots \wedge \omega^{i_r}.$$

Permuting the order of $i_1, \ldots, i_r$ leaves the right side unchanged, except for a possible change of sign according to Lemma 6.9. It follows that the set of $\binom{n}{r}$ elements of the form $\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$ with $1 \leq i_1 < i_2 < \cdots < i_r \leq n$ span $\bigwedge^r(V)$. On the other hand, they are independent. For if we suppose that some linear combination of them is zero, say

$$\sum_{i_1 < \cdots < i_r} \alpha_{i_1 \cdots i_r} \omega^{i_1} \wedge \cdots \wedge \omega^{i_r} = 0,$$

then its value on each set of r basis vectors must be zero. In particular, given $k_1 < \cdots < k_r$, we have

$$0 = (\sum \alpha_{i_1 \cdots i_r} \omega^{i_1} \wedge \cdots \wedge \omega^{i_r})(\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_r}),$$

which becomes $\alpha_{k_1 \cdots k_r} = 0$ by virtue of the formula of Corollary 6.7 combined with the fact that $\omega^i(\mathbf{e}_k) = \delta^i_k$ for $1 \leq i, k \leq n$. By suitable choice of $k_1 < \cdots < k_r$, we see that each coefficient must be zero; therefore the given set of elements of $\bigwedge^r(V)$ is linearly independent and a basis.

To complete the proof we note that

$$\dim \bigwedge(V) = \sum_{r=0}^{n} \dim \bigwedge^r(V) = \sum_{r=0}^{n} \binom{n}{r} = 2^n. \qquad \blacksquare$$

The following theorem is an immediate consequence of Theorem 6.2, the fact that $\mathscr{A} \circ F^* = F^* \circ \mathscr{A}$, and the definition of exterior multiplication.

(6.11) Theorem *Let V and W be finite-dimensional vector spaces and $F_*: W \to V$ a linear mapping. Then $F^*: \mathscr{T}(V) \to \mathscr{T}(W)$ takes $\bigwedge(V)$ into $\bigwedge(W)$ and is a homomorphism of these (exterior) algebras.*

The Exterior Algebra on Manifolds

It is evident from what we have seen above that all of these ideas extend to alternating tensor fields on a C^∞ manifold M. We introduce the following terminology:

(6.12) Definition An alternating covariant tensor field of order r on M will be called an *exterior differential form of degree r* (or sometimes simply *r-form*).

The set $\bigwedge^r(M)$ of all such forms is a subspace of $\mathscr{T}^r(M)$. The following two theorems are immediate consequences of what has been done above and their proofs will be left to the reader. We let M, N be manifolds and $F: M \to N$ be a C^∞ mapping.

(6.13) Theorem *Let $\bigwedge(M)$ denote the vector space over $\boldsymbol{R}$ of all exterior differential forms. Then for $\varphi \in \bigwedge^r(M)$ and $\psi \in \bigwedge^s(M)$ the formula $(\varphi \wedge \psi)_p = \varphi_p \wedge \psi_p$ defines an associative product satisfying $\varphi \wedge \psi = (-1)^{rs}\psi \wedge \varphi$. With this product, $\bigwedge(M)$ is an algebra over $\boldsymbol{R}$. If $f \in C^\infty(M)$, we also have $(f\varphi) \wedge \psi = f(\varphi \wedge \psi) = \varphi \wedge (f\psi)$. If $\omega^1, \ldots, \omega^n$ is a field of coframes on M (or an open set U of M), then the set $\{\omega^{i_1} \wedge \cdots \wedge \omega^{i_r} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n\}$ is a basis of $\bigwedge^r(M)$ (or $\bigwedge^r(U)$, respectively).*

(6.14) Theorem *If $F: M \to N$ is a C^∞ mapping of manifolds, then $F^*: \bigwedge(N) \to \bigwedge(M)$ is an algebra homomorphism.*

We shall call $\bigwedge(M)$ the *algebra of differential forms* or *exterior algebra* on M.

Exercises

1. By constructing an example show that tensor products are not commutative and that φ, ψ symmetric (respectively alternating) does not imply $\varphi \otimes \psi$ is symmetric (respectively, alternating).
2. Let $\Sigma(V)$ denote the subspace of $\mathscr{T}(V)$ consisting of all $\varphi \in \mathscr{T}(V)$ whose constituents are symmetric: $\Sigma(V)$ is the image of $\Sigma^0(V) \oplus \Sigma^1(V) \oplus \cdots \oplus \Sigma^r(V) \oplus \cdots$ in $\mathscr{T}(V)$ under the natural injection defined by inclusion of $\Sigma^r(V)$ in $\mathscr{T}^r(V)$. Define an associative multiplication in $\Sigma(V)$ by analogy with that for $\bigwedge(V)$ and prove analogs of the theorems proved for $\bigwedge(V)$ where possible.
3. If $\varphi_i \in \bigwedge^1(V)$, $i = 1, \ldots, k$, show that $\varphi_1 \wedge \cdots \wedge \varphi_k(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \det(\varphi_i(\mathbf{v}_j))$, a $k \times k$ determinant.
4. Show that the C^∞ exterior forms of order r on M are exactly the functions $\Phi: \mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M) \to C^\infty(M)$ which are multilinear, in the sense of $C^\infty(M)$ modules, and alternating. Find an example of such a function which is $\boldsymbol{R}$-linear but not a tensor field. (*Hint:* Use $[X, Y]$.)
5. Let $\varphi_1, \ldots, \varphi_r$ be elements of $V^* = \bigwedge^1(V)$. Show that they are linearly dependent if and only if $\varphi_1 \wedge \cdots \wedge \varphi_r = 0$.
6. Assume $\varphi \in \bigwedge^r(V)$ and $\mathbf{v} \in V$. Define an element $i(\mathbf{v})\varphi$ of $\bigwedge^{r-1}(V)$ by
$$(i(\mathbf{v})\varphi)(\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}) = \varphi(\mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_{r-1}).$$
Show that $i(\mathbf{v})$ thus defined determines a linear mapping of $\bigwedge^r(V)$ into $\bigwedge^{r-1}(V)$ and that if $\varphi \in \bigwedge^r(V)$, $\psi \in \bigwedge^s(V)$, then $i(\mathbf{v})(\varphi \wedge \psi) = (i(\mathbf{v})\varphi) \wedge \psi + (-1)^r \varphi \wedge (i(\mathbf{v})\psi)$. Extend this definition and these properties to exterior forms on a manifold (with $\mathbf{v}$ replaced by a vector field).
7. A Riemannian metric Φ on a manifold is often denoted by ds^2 in local coordinates $x^1, \ldots, x^n$ on M with $ds^2 = \sum_{i,j=1}^n g_{ij}(x)\, dx^i\, dx^j$. Interpret this by the use of tensor multiplication and Theorem 6.3. Show that

from this expression for Φ and the formulas $dx^i = \sum_{j=1}^n (\partial x^i/\partial y^j)\, dy^j$ we may derive the formula for change of components of Φ relative to a change of local coordinates.

7 Orientation of Manifolds and the Volume Element

We shall make one application of differential forms in this paragraph, others in subsequent chapters. To do this we shall need the concept of an oriented vector space. Let V be a vector space and $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ be bases. The bases are said to have the *same orientation* if the determinant of the matrix of coefficients expressing one basis in terms of the other is positive, that is, if $\det(\alpha_i^j) > 0$, where $\mathbf{f}_i = \sum_{j=1}^n \alpha_i^j \mathbf{e}_j$, $i = 1, \ldots, n$. The reader should check that this is an equivalence relation on the set of all bases (or frames) of V and that there are exactly two equivalence classes. A choice of one of these is said to orient V so that we have the following definition:

(7.1) Definition An *oriented vector space* is a vector space plus an equivalence class of allowable bases: all those bases with the same orientation as a chosen one; they will be called *oriented* or *positively oriented* bases or frames.

This concept is related to the choice of a basis Ω of $\bigwedge^n(V)$. Recall that $\dim \bigwedge^n(V) = \binom{n}{n} = 1$, so that any nonzero element is a basis. The relationship to orientation appears as a corollary to the following lemma:

(7.2) Lemma *Let $\Omega \neq 0$ be an alternating covariant tensor on V of order $n = \dim V$ and let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis of V. Then for any set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with $\mathbf{v}_i = \sum \alpha_i^j \mathbf{e}_j$, we have*

$$\Omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \det(\alpha_i^j)\Omega(\mathbf{e}_1, \ldots, \mathbf{e}_n).$$

Proof This lemma says that up to a nonvanishing scalar multiple Ω is the determinant of the components of its variables. In particular, if $V = V^n$ is the space of n-tuples and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the canonical basis, then $\Omega(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is proportional to the determinant whose rows are $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The proof is a consequence of the definition of determinant. Given Ω and $\mathbf{v}_1, \ldots, \mathbf{v}_n$, we use the linearity and antisymmetry of Ω to write

$$\Omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{j_1, \ldots, j_n} \alpha_1^{j_1} \cdots \alpha_n^{j_n} \Omega(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_n})$$

Since $\Omega(e_{j_1}, \ldots, e_{j_n}) = 0$ if two indices are equal, we may write

$$\Omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn} \sigma \alpha_1^{\sigma(1)} \cdots \alpha_n^{\sigma(n)} \Omega(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \det(\alpha_i^j)\Omega(\mathbf{e}_1, \ldots, \mathbf{e}_n).$$

The last equality is the standard definition of determinant ($\mathfrak{S}_n$ is the symmetric group on n letters). ∎

(7.3) Corollary *A nonvanishing $\Omega \in \bigwedge^n(V)$ has the same sign (or opposite sign) on two bases if they have the same (respectively, opposite) orientation; thus choice of an $\Omega \neq 0$ determines an orientation of V. Two such forms Ω_1, Ω_2 determine the same orientation if and only if $\Omega_1 = \lambda\Omega_2$, where λ is a positive real number.*

Proof From the formula of the lemma we see that Ω has the same sign on equivalent bases and opposite sign on inequivalent bases. If $\lambda > 0$, then $\lambda\Omega$ has the same sign on any basis as Ω does, whereas the contrary holds if $\lambda < 0$. ∎

(7.4) Remark Note that if $\Omega \neq 0$, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if and only if $\Omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) \neq 0$. Also note that the formula of the lemma can be construed as a formula for change of component of Ω—there is just *one* component since $\dim \bigwedge^n(V) = 1$—when we change from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. These statements are immediate consequences of the formula in the lemma.

If V is a Euclidean vector space, that is, has a positive definite inner product $\Phi(\mathbf{v}, \mathbf{w})$, then in orienting V we may choose an orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to determine the orientation and choose an n-form Ω whose value on $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is $+1$. If $\mathbf{f}_i = \sum \alpha_i^j \mathbf{e}_j$ is another orthonormal basis, then

$$\Omega(\mathbf{f}_1, \ldots, \mathbf{f}_n) = \det(\alpha_i^j)\Omega(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \pm 1,$$

depending on whether $\mathbf{f}_1, \ldots, \mathbf{f}_n$ is similarly or oppositely oriented. [We have used the fact that the determinant of an orthogonal matrix is ± 1.] Thus the value of Ω on any orthonormal basis is ± 1 and Ω is uniquely determined up to its sign by this property. In this case the form Ω may be given a geometric meaning when $n = 2$ or 3; $\Omega(\mathbf{v}_1, \mathbf{v}_2)$ or $\Omega(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is the area or volume, respectively, of the parallelogram or parallelepiped of which the given vectors are the sides from the origin. This is a standard formula from analytical geometry and serves as a geometric motivation for some later applications. (See Exercise 2.)

To extend the concept of orientation to a manifold M one must try to orient each of the tangent spaces $T_p(M)$ in such a way that orientation of nearby tangent spaces agree. We will do this in two ways and then demonstrate their equivalence as an application of the ideas of this chapter.

(7.5) Definition We shall say that M is *orientable* if it is possible to define a C^∞ n-form Ω on M which is not zero at any point—in which case M is said to be *oriented* by the choice of Ω.

By virtue of Corollary 7.3 any such Ω orients each tangent space. Of course any form $\Omega' = \lambda\Omega$, where $\lambda > 0$ is a C^∞ function, would give M the same orientation.

Thus $\boldsymbol{R}^n$ with the form $\Omega = dx^1 \wedge \cdots \wedge dx^n$ is an example; this is known as the *natural orientation* of $\boldsymbol{R}^n$ and corresponds to the orientation of the frames $\partial/\partial x^1, \ldots, \partial/\partial x^n$. If $U \subset \boldsymbol{R}^n$ is an open set, it is oriented by $\Omega_U = \Omega|_U$ and we say that a diffeomorphism $F\colon U \to V \subset \boldsymbol{R}^n$ is orientation preserving if $F^*\Omega_V = \lambda\Omega_U$, $\lambda > 0$ a C^∞ function on U. More generally a diffeomorphism $F\colon M_1 \to M_2$ of manifolds oriented by Ω_1, Ω_2, respectively, is *orientation-preserving* if $F^*\Omega_2 = \lambda\Omega_1$, where $\lambda > 0$ is a C^∞ function on M.

A second, perhaps more natural definition of orientability could be given as follows: M is orientable if it can be covered with *coherently oriented* coordinate neighborhoods $\{U_\alpha, \varphi_\alpha\}$, that is, neighborhoods such that if $U_\alpha \cap U_\beta \neq \varnothing$, then $\varphi_\alpha \circ \varphi_\beta^{-1}$ is orientation-preserving. We shall now see that this second definition is equivalent to Definition 7.5.

(7.6) Theorem *A manifold M is orientable if and only if it has a covering $\{U_\alpha, \varphi_\alpha\}$ of coherently oriented coordinate neighborhoods.*

Proof First suppose that M is orientable and let Ω be a nowhere vanishing n-form that determines the orientation. We choose any covering $\{U_\alpha, \varphi_\alpha\}$ by coordinate neighborhoods, with local coordinates $x_\alpha^1, \ldots, x_\alpha^n$ such that for Ω restricted to U_α we have the expression in local coordinates

$$\varphi_\alpha^{-1*}\Omega_{U_\alpha} = \lambda_\alpha(x)\, dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n \qquad \text{with} \quad \lambda_\alpha > 0.$$

We may easily choose coordinates so that the scalar function λ_α, component of Ω, is positive on U_α, since replacing coordinates $(x^1, \ldots, x^n)$ by $(-x^1, \ldots, x^n)$, that is, changing the sign of one coordinate, changes the sign of λ. An easy computation, using Lemma 7.2 and Remark 7.4, shows that if $U_\alpha \cap U_\beta \neq \varnothing$, then on this set the formula for change of component is

$$\text{(7.6}'\text{)} \qquad \lambda_\alpha \det\left(\frac{\partial x_\alpha^i}{\partial x_\beta^j}\right) = \lambda_\beta\,.$$

Since $\lambda_\alpha > 0$ and $\lambda_\beta > 0$, the determinant of the Jacobian is positive, so the coordinate neighborhoods we have chosen are coherently oriented.

Now suppose that M has a covering by coherently oriented coordinate neighborhoods $\{U_\alpha, \varphi_\alpha\}$. We use a subordinate partition of unity $\{f_i\}$ to construct an n-form Ω on M which does not vanish at any point. For each $i = 1, 2, \ldots$ we choose a coordinate neighborhood $U_{\alpha_i}, \varphi_{\alpha_i}$ of the covering such that $U_{\alpha_i} \supset \operatorname{supp} f_i$; these neighborhoods, which we relabel U_i, φ_i, cover M. If $U_i \cap U_j \neq \varnothing$, then by assumption the determinant of the Jacobian matrix of $\varphi_i \circ \varphi_j^{-1}$ is positive on $U_i \cap U_j$. Define $\Omega \in \bigwedge^n(M)$ by

$$\Omega = \sum_i f_i \varphi_i^*(dx_i^1 \wedge \cdots \wedge dx_i^n),$$

extending each summand to all of M by defining it to be zero outside the closed set $\operatorname{supp} f_i$. Letting $p \in M$ be arbitrary, we will show $\Omega_p \neq 0$. By the

local finiteness of $\{\operatorname{supp} f_i\}$ we may choose a coordinate neighborhood V, ψ of p which is coherently oriented to the U_i, φ_i and intersects only a finite number of the sets $\operatorname{supp} f_i$, say for $i = i_1, \ldots, i_k$. Let $y^1, \ldots, y^n$ be the local coordinates in V and use formula (7.6′), on each summand to change components:

$$\Omega_p = \sum_{j=1}^{k} f_{i_j}(p)\varphi_{i_j}^*(dx_{i_j}^1 \wedge \cdots \wedge dx_{i_j}^n) = \sum f_{i_j}(p) \det\left(\frac{\partial x_{i_j}^k}{\partial y^l}\right)_{\psi(p)} \psi^*(dy^1 \wedge \cdots \wedge dy^n).$$

Now each $f_{i_j} \geq 0$ on M and at least one of them is positive at p; moreover, the Jacobian determinants are all positive. This implies $\Omega_p \neq 0$ and since p was arbitrary, Ω is never zero on M. ∎

A Riemannian manifold has the special property that the tangent space $T_p(M)$ at every point p has an inner product. Applying our remarks about n-forms on a Euclidean vector space of dimension n, we have the following theorem:

(7.7) Theorem *Let M be an orientable Riemannian manifold with Riemannian metric Φ. Corresponding to an orientation of M there is a uniquely determined n-form Ω which gives the orientation and which has the value $+1$ on every oriented orthonormal frame.*

Proof It is clear from our earlier discussion that at each point $p \in M$, Ω_p is determined uniquely by the requirement that on any oriented orthonormal basis $F_{1p}, \ldots, F_{np}$ of $T_p(M)$ we have $\Omega_p(F_{1p}, \ldots, F_{np}) = +1$. Let U, φ be any coordinate neighborhood with coordinate frames $E_1, \ldots, E_n$. The functions $g_{ij}(p) = \Phi_p(E_{ip}, E_{jp})$, $p \in U$, define the components of Φ relative to these local coordinates and are C^∞ by definition. We shall derive an expression for the component $\Omega(E_1, \ldots, E_n)$ on U in terms of the matrix (g_{ij}), from which it will be apparent that Ω is a C^∞ n-form. Choose at $p \in U$ any oriented, orthonormal basis $F_{1p}, \ldots, F_{np}$ and let the $n \times n$ matrix (α_i^k) denote the components of $E_{1p}, \ldots, E_{np}$ with respect to this basis:

$$E_{ip} = \sum_{k=1}^{n} \alpha_i^k F_{kp}, \qquad i = 1, \ldots, n.$$

Then since $\Phi(F_{kp}, F_{lp}) = \delta_{kl}$, we have

$$g_{ij}(p) = \Phi_p(E_{ip}, E_{jp}) = \Phi_p\left(\sum_k \alpha_i^k F_{kp}, \sum_l \alpha_j^l F_{lp}\right) = \sum_{k=1}^{n} \alpha_i^k \alpha_j^k$$

for $1 \leq i, j \leq n$. This may be written as a matrix equation:

$$(g_{ij}(p)) = {}^t\!AA,$$

the product of the transpose of $A = (\alpha_i^k)$ with A itself.

On the other hand $\Omega_p(E_{1p}, \ldots, E_{np}) = \det(\alpha_i^k)\Omega_p(F_{1p}, \ldots, F_{np})$ by Lemma 7.2, and $\Omega(F_{1p}, \ldots, F_{np}) = +1$ by our definitions. Since $\det({}^tAA) = (\det A)^2 = \det(g_{ij})$, this gives for the component of Ω in local coordinates

$$\Omega_p(E_{1p}, \ldots, E_{np}) = \left(\det(g_{ij}(p))\right)^{1/2},$$

which tells us that the component, being the square root of a positive C^∞ function of $p \in U$, is itself a C^∞ function on the local coordinate neighborhood U. Since U, φ is arbitrary, Ω is a C^∞ n-form on M. ∎

This form Ω is called the (natural) *volume element* of the oriented Riemannian manifold. We have just seen that in local coordinates we have the following expression for Ω:

(7.8) $$\varphi^{-1*}\Omega = \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n,$$

where $g(x) = \det(g_{ij}(x))$. [We use the same notation for g_{ij} as functions on U and on $\varphi(U)$.] When $M = \boldsymbol{R}^n$ with the usual coordinates and metric, this becomes $\Omega = dx^1 \wedge \cdots \wedge dx^n$. In this case, as we remarked earlier, the value of Ω_p on a set of vectors is the volume of the parallelepiped whose edges from p are these vectors. In particular, on the unit cube with vertex at p and sides $\partial/\partial x^1, \ldots, \partial/\partial x^n$, Ω has the value $+1$. As might be anticipated, the existence of the form Ω on a Riemannian manifold will enable us to define the volume of suitable subsets of the manifold and to extend to these manifolds the volume integrals defined in $\boldsymbol{R}^n$ in integral calculus.

Exercises

1. Using the definition of $dx^1 \wedge \cdots \wedge dx^n$ on $\boldsymbol{R}^n$ from Corollary 6.7, show that its value on $\partial/\partial x^1, \ldots, \partial/\partial x^n$ is indeed $+1$ so that this is Ω for $\boldsymbol{R}^n$ with the standard Riemannian metric, as claimed above.
2. Prove that the volume of the parallelepiped of $\boldsymbol{R}^3$ whose vertex is at the origin and whose sides (from this vertex) are the vectors $\mathbf{v}_i = (x_i^1, x_i^2, x_i^3)$, $i = 1, 2, 3$, is in fact the determinant of the matrix (x_i^j).
3. Show that $n \times n$ determinants as functions of the n-rows $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are completely characterized by being alternating n-tensors on V^n whose value on the standard basis is $+1$.
4. Compute the expression for Ω on S^2 (with the induced metric of $\boldsymbol{R}^3$) in terms of the coordinates given by: (i) stereographic projection and (ii) spherical coordinates (ρ, θ, φ) with $\rho = 1$.

8 Exterior Differentiation

Much of this chapter has been devoted to extending the concepts of covariant tensors, and of operations on covariant tensors, on a single vector space V to tensors and tensor operations on manifolds. This was done

according to a very standardized procedure which consisted in viewing each tangent space $T_p(M)$ as a copy of V and thus extending the tensor or tensor operation point-by-point, making suitable restrictions to ensure some sort of smooth variation. By using very different ideas we now introduce an important operator d mapping $\bigwedge(M)$ onto itself. It is defined in terms of differentiation and is known as the *exterior derivative*; it has no analog on $\bigwedge(V)$, the exterior algebra of a single vector space.

When U is an open subset of M we shall denote by θ_U the restriction of an exterior form on M to U; of course $\theta_U = i^*\theta$, $i: U \to M$ being the inclusion map. When U, φ is a coordinate neighborhood with $x^1, \ldots, x^n$ as coordinate functions on U, that is, $\varphi(q) = (x^1(q), \ldots, x^n(q))$, then the differentials of these functions $dx^1, \ldots, dx^n$ are linearly independent elements of $\bigwedge^1(U)$ and constitute a C^∞ field of coframes on U. It follows that they, with 1, generate $\bigwedge(U)$ over $C^\infty(U)$, or equivalently, $C^\infty(U) = \bigwedge^0(U)$ and $\bigwedge^1(U)$ generate the algebra $\bigwedge(U)$ over $\boldsymbol{R}$. Thus *locally* every k-form θ on M has a unique representation on U of the form

$$\theta_U = \sum_{i_1 < \cdots < i_k} a_{i_1 \cdots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \qquad a_{i_1 \cdots i_k} \in C^\infty(U),$$

the summation being over all sets of indices such that $1 \le i_1 < i_2 < \cdots < i_k \le n$. If we define $b_{i_1 \cdots i_k}$ for all values of the indices so as to change sign whenever two indices are permuted—in particular to be zero if two indices are equal—and to equal $a_{i_1 \cdots i_k}$ if $i_1 < \cdots < i_k$, then we also have a unique representation

$$\theta_U = \sum \frac{1}{k!} b_{i_1 \cdots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

the summation being over all values of the indices. Both representations are used in practice. We are using $dx^1, \ldots, dx^n$ to denote the coordinate coframes, rather than $\omega^1, \ldots, \omega^n$ as in Section 1, in order to emphasize that the dx^i are differentials of *functions* on $U \subset M$. This is important in what follows.

(8.1) Theorem *Let M be any C^∞ manifold and let $\bigwedge(M)$ be the algebra of exterior differential forms on M. Then there exists a unique $\boldsymbol{R}$-linear map $d_M\colon \bigwedge(M) \to \bigwedge(M)$ such that:*

(1) *if $f \in \bigwedge^0(M) = C^\infty(M)$, then $d_M f = df$, the differential of f;*
(2) *if $\theta \in \bigwedge^r(M)$ and $\sigma \in \bigwedge^s(M)$, then $d_M(\theta \wedge \sigma) = d_M \theta \wedge \sigma + (-1)^r \theta \wedge d_M \sigma$;*
(3) $d_M^2 = 0$.

This map will commute with restriction to open sets $U \subset M$, i.e., $(d_M \theta)_U = d_U \theta_U$, and map $\bigwedge^r(M)$ into $\bigwedge^{r+1}(M)$.

Proof We give the proof in a series of steps.

(A) We remark that if d_M exists and $g, f^1, \ldots, f^r \in C^\infty(M)$, then (1)–(3) imply that for $\theta = g\, df^1 \wedge \cdots \wedge df^r$ we must have

$$d_M \theta = dg \wedge df^1 \wedge \cdots \wedge df^r.$$

Now suppose that M is covered by a single coordinate neighborhood U, φ with coordinate functions $x^1, \ldots, x^n$. The above remark and linearity imply that d_M must be given by the formula

$$(*) \qquad d_M\left(\sum a_{i_1 \cdots i_r}\, dx^{i_1} \wedge \cdots \wedge dx^{i_r}\right) = \sum da_{i_1 \cdots i_r} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_r},$$

where

$$da_{i_1 \cdots i_r} = \sum_{j=1}^{n} \frac{\partial a_{i_1 \cdots i_r}}{\partial x^j}\, dx^j$$

and the summation in $(*)$ is over $1 \le i_1 < i_2 < \cdots < i_r \le n$. Therefore, if defined at all, d_M is unique in this case.

Conversely, the d_M defined by $(*)$ is linear and trivially satisfies (1) and (3). To check (2) it is enough to consider forms $\theta = a\, dx^{i_1} \wedge \cdots \wedge dx^{i_r}$ and $\sigma = b\, dx^{j_1} \wedge \cdots \wedge dx^{j_s}$, the general statement being then a consequence of linearity. We have

$$\begin{aligned}
&d_M[(a\, dx^{i_1} \wedge \cdots \wedge dx^{i_r}) \wedge (b\, dx^{j_1} \wedge \cdots \wedge dx^{j_s})] \\
&\quad = d_M(ab)(dx^{i_1} \wedge \cdots \wedge dx^{i_r}) \wedge (dx^{j_1} \wedge \cdots \wedge dx^{j_s}) \\
&\quad = [(d_M a)b + a(d_M b)] \wedge (dx^{i_1} \wedge \cdots \wedge dx^{i_s}) \wedge (dx^{j_1} \wedge \cdots \wedge dx^{j_s}) \\
&\quad = (d_M a \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_r}) \wedge (b\, dx^{j_1} \wedge \cdots \wedge dx^{j_s}) \\
&\qquad + (-1)^r (a\, dx^{i_1} \wedge \cdots \wedge dx^{i_r}) \wedge (db \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_s}),
\end{aligned}$$

which completes the proof. The $(-1)^r$ is due to the fact that

$$db \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_r} = (-1)^r\, dx^{i_1} \wedge \cdots \wedge dx^{i_r} \wedge db.$$

(B) Now suppose that $d_M\colon \bigwedge(M) \to \bigwedge(M)$ with properties (1)–(3) is defined and that $U \subset M$ is a coordinate neighborhood on M with coordinate functions $x^1, \ldots, x^n$. According to (A), $d_U\colon \bigwedge(U) \to \bigwedge(U)$ is uniquely defined by $(*)$. We will show that for any $\theta \in \bigwedge(M)$ the restriction of $d_M \theta$ to U is equal to d_U applied to θ restricted to U:

$$(d_M \theta)_U = d_U \theta_U\,.$$

We may suppose that $\theta \in \bigwedge^r(M)$ and that

$$\theta_U = \sum a_{i_1 \cdots i_r}\, dx^{i_1} \wedge \cdots \wedge dx^{i_r}, \qquad a_{i_1 \cdots i_r} \in C^\infty(U).$$

Suppose p is an arbitrary point of U. Applying Corollary III.3.5 to an open set W, $p \in W$ and $\overline{W} \in U$, we may find a neighborhood V of p with $V \subset W$ and C^∞ functions $y^1, \ldots, y^n$ and $b_{i_1 \cdots i_r}$ on M vanishing outside W but are identical to $x^1, \ldots, x^n$ and $a_{i_1 \cdots i_r}$, respectively, on V. Define $\sigma \in \bigwedge^r(M)$ by

$$\sigma = \sum b_{i_1 \cdots i_r}\, dy^{i_1} \wedge \cdots \wedge dy^{i_r}.$$

Then σ is an r-form on M which vanishes outside W and is identical to θ on V. Now let g be a C^∞ function on M which has the value $+1$ at p and is zero outside V. The r-form $g(\theta - \sigma)$ vanishes everywhere on M as does $dg \wedge (\theta - \sigma)$. Therefore, using (A),

$$g\, d_M \theta = g\, d_M \sigma = g \sum da_{i_1 \cdots i_r} \wedge dy^{i_1} \wedge \cdots \wedge dy^{i_r}.$$

On V we have

$$\sum db_{i_1 \cdots i_r} \wedge dy^{i_1} \wedge \cdots \wedge dy^{i_r} = \sum da_{i_1 \cdots i_r} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_r}$$

so that at the point p, where $g(p) = 1$, $d_M \theta = d_U \theta_U$. Since p is arbitrary, this holds throughout U.

(C) If $d_M \colon \bigwedge(M) \to \bigwedge(M)$ satisfying (1)–(3) exists, it is unique. Indeed, let $\{U_\alpha, \varphi_\alpha\}$ be any covering of M by coordinate neighborhoods; each d_{U_α} exists by (A); and for any $\theta \in \bigwedge(M)$ we have $(d_M \theta)_{U_\alpha} = d_{U_\alpha} \theta_{U_\alpha}$ for any U_α by (B). Since every $p \in M$ lies in a neighborhood U_α, this would determine d_M completely.

On the other hand, we may use this formula to define d_M. To do so we must verify that if $p \in U_\alpha \cap U_\beta$, then $d_M \theta$ is uniquely determined at p. This essentially repeats the argument above: Let $U = U_\alpha \cap U_\beta$; applying (A) and (B) to U, an open subset and coordinate neighborhood with coordinate map φ_β cut down to U, we have

$$(d_{U_\alpha} \theta_{U_\alpha})_U = d_U \theta_U = (d_{U_\beta} \theta_{U_\beta})_U\,.$$

Therefore $(d_M \theta)_{U_\alpha}$ is determined on every U_α in such a manner that $(d_M \theta)_{U_\alpha} = (d_M \theta)_{U_\beta}$ on points common to U_α and U_β. This determines d_M.

Because (1)–(3) hold on each U_α and the other operations of exterior algebra commute with restriction: $(\theta + \sigma)_U = \theta_U + \sigma_U$, $(\theta \wedge \sigma)_U = \theta_U \wedge \sigma_U$, etc., d_M has the required properties as an operator on $\bigwedge(M)$. This completes the proof. ∎

Since d_M is uniquely defined for every C^∞ manifold M, we can drop the subscript M and use d to denote all of these operators. We know from the above proof that d commutes with restriction of differential forms to coordinate neighborhoods. It is important to know how it behaves relative to a C^∞ mapping $F \colon M \to N$. Any such mapping, as we know, induces a homomorphism $F^* \colon \bigwedge(N) \to \bigwedge(M)$. The following theorem gives the relation of F^* and d.

(8.2) Theorem *F^* and d commute, that is $F^* \circ d = d \circ F^*$.*

Proof We know that both F^* and d are $\boldsymbol{R}$-linear and that the equality $F^*(d\varphi) = d(F^*\varphi)$ holds on M if it holds locally. More precisely, by virtue of the facts concerning d determined above we see that the theorem will hold if we can establish it for any pairs V, ψ, U, θ of coordinate neighborhoods on M, N, respectively, such that $F(V) \subset U$. Let $m = \dim M$ and $n = \dim N$ and $x^1, \ldots, x^m$ and $y^1, \ldots, y^n$ be the coordinate functions on V, U, respectively, with $y^j = y^j(x^1, \ldots, x^m)$, $j = 1, \ldots, n$, giving the map F in local coordinates. Then it is enough to establish $F^* \circ d = d \circ F^*$ on forms of the following type:

$$\varphi = a(y)\, dy^{i_1} \wedge \cdots \wedge dy^{i_k},$$

any other forms being the sum of such. We proceed by induction on the degree of the forms. For forms $a(y)$ of degree zero, that is, C^∞ functions, we have

$$F^*(da) = d(F^*a),$$

since

$$F^*(da)(X_p) = da(F_* X_p) = (F_* X_p)a = X_p(a \circ F) = X_p(F^*a) = d(F^*a)(X_p).$$

(*Note:* By definition, $F^*a = a \circ F$.)

Suppose the theorem to be true for all forms of degree less than k and let φ be a k-form of the type above. Let $\varphi_1 = a\, dy^{i_1}$ and $\varphi_2 = dy^{i_2} \wedge \cdots \wedge dy^{i_k}$ so that $\varphi = \varphi_1 \wedge \varphi_2$ with both φ_1 and φ_2 of degree less than k; moreover since $d^2 = 0$, we have $d\varphi_2 = 0$. Thus

$$\begin{aligned} d(F^*(\varphi_1 \wedge \varphi_2)) &= d[(F^*\varphi_1) \wedge (F^*\varphi_2)] \\ &= (dF^*\varphi_1) \wedge (F^*\varphi_2) - (F^*\varphi_1) \wedge (dF^*\varphi_2) \\ &= F^*(d\varphi_1) \wedge F^*\varphi_2 = F^*(d\varphi_1 \wedge \varphi_2) = F^*d(\varphi_1 \wedge \varphi_2). \quad \blacksquare \end{aligned}$$

An Application to Frobenius' Theorem

The algebra of exterior differential forms $\bigwedge(M)$ on a C^∞ manifold M, with the operator d just defined, is very important in the application of calculus to manifolds. Forms are involved in integration on manifolds (especially in extending Gauss', Stokes' and Green's theorems); in the algebraic topology of the manifold via the theorems of de Rham and Hodge; and in the study of partial differential equations. We will touch on the first two topics later. As to the differential equations aspect, we will show next that the essential data and hypothesis of Frobenius' theorem can be stated in terms of d and $\bigwedge(M)$.

On a vector space $\boldsymbol{V}$ of dimension n, a k-dimensional subspace $\boldsymbol{D}$ may be determined in either of two equivalent ways: (i) by giving a basis $\mathbf{e}_1, \ldots, \mathbf{e}_k$ of

$\boldsymbol{D}$ or (ii) by giving $n - k$ linearly independent elements of $\boldsymbol{V}^*$, say $\varphi^{k+1}, \ldots, \varphi^n$ which are zero on $\boldsymbol{D}$. In fact we may extend $\mathbf{e}_1, \ldots, \mathbf{e}_k$ to a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\boldsymbol{V}$ so that $\varphi^{k+1}, \ldots, \varphi^n$ is part of a dual basis $\varphi^1, \ldots, \varphi^n$ of $\boldsymbol{V}^*$.

Similarly, if Δ is a C^∞ distribution of dimension k on M, an n-manifold, then locally, say in a coordinate neighborhood V, ψ, we may suppose Δ is defined by $n - k$ linearly independent 1-forms $\varphi^{k+1}, \ldots, \varphi^n$. We may restate the condition that Δ be involutive—hence Frobenius' theorem—as follows:

(8.3) Theorem *Let Δ be a C^∞ distribution of dimension k on M, $\dim M = n$. Then Δ is involutive if and only in a neighborhood V of each $p \in M$ there exist $n - k$ linearly independent one-forms $\varphi^{k+1}, \varphi^{k+2}, \ldots, \varphi^n$ which vanish on Δ and satisfy the condition*

$$d\varphi^r = \sum_{l=k+1}^{n} \theta_l^r \wedge \varphi^l, \qquad r = k+1, \ldots, n,$$

for suitable 1-forms θ_l^r.

Proof This may be considered a sort of dual statement to our earlier condition on Δ in terms of the existence of a local basis $X_1, \ldots, X_k$ at each point. (Just as in that case, we may state the conditions in a fashion which does not depend on local bases. This will be done below (8.7), with proof left to the exercises.)

Suppose a distribution Δ is given. Then in a neighborhood V of each point a local basis $X_1, \ldots, X_k$ of Δ can be completed to a field of frames $X_1, \ldots, X_k, \ldots, X_n$. If $\varphi^1, \ldots, \varphi^k, \varphi^{k+1}, \ldots, \varphi^n$ is the uniquely determined dual field of coframes, then $\varphi^{k+1}, \ldots, \varphi^n$ vanish on $X_1, \ldots, X_k$ and hence on Δ. The distribution Δ is involutive by Definition IV.8.2 if and only if in the expressions $[X_i, X_j] = \sum_{l=1}^n c_{ij}^l X_l$, giving $[X_i, X_j]$ as linear combinations of the basis, we have $c_{ij}^l = 0$ for $1 \le i, j \le k$ and $k + 1 \le l \le n$.

(8.4) Lemma *Let $\omega \in \bigwedge^1(M)$ and $X, Y \in \mathfrak{X}(M)$. Then we have*

$$d\omega(X, Y) = X\omega(Y) - Y\omega(X) - \omega([X, Y]).$$

Let us assume the lemma and proceed with the proof. We compute $d\varphi^r$, using the lemma and recalling that $\varphi^i(X_j)$ is constant for $1 \le i, j \le n$. We have then,

$$d\varphi^r(X_i, X_j) = -\varphi^r([X_i, X_j]) = -\sum_{l=1}^{n} c_{ij}^l \varphi^r(X_l) = -c_{ij}^r$$

for $1 \le i, j, r \le n$. On the other hand

$$d\varphi^r = \tfrac{1}{2} \sum_{s,t}^{n} b_{st}^r \varphi^s \wedge \varphi^t, \qquad 1 \le r \le n,$$

where b_{st}^r are uniquely determined if we assume $b_{st}^r = -b_{ts}^r$. Hence

$$d\varphi^r(X_i, X_j) = \tfrac{1}{2} \sum_{s,t} b_{st}^r[\varphi^s(X_i)\varphi^t(X_j) - \varphi^t(X_i)\varphi^s(X_j)]$$
$$= \tfrac{1}{2}(b_{ij}^r - b_{ji}^r) = b_{ij}^r.$$

From this we have $b_{ij}^r = -c_{ij}^r$ and so the system is involutive if and only if for each $r > k$

$$d\varphi^r = \sum_{l=k+1}^{n} \left\{ \sum_{i=1}^{k} b_{il}^r \varphi^i + \sum_{j=k+1}^{n} \tfrac{1}{2} b_{jl}^r \varphi^j \right\} \wedge \varphi^l,$$

that is, the terms involving b_{ij}^r with $1 \le i, j \le k$ and $r > k$ vanish. Taking the terms in $\{\ \}$ as θ_l^r we have completed the proof except for Lemma 8.4. ∎

Proof of Lemma 8.4 It is enough to prove that it is true locally, say in a coordinate neighborhood of each point. In any such neighborhood with coordinates $x^1, \ldots, x^n$, $\omega = \sum_{i=1}^n a_i\, dx^i$ and it is easy to see that the equation of the lemma holds for all ω if it holds for every ω of the form $f\,dg$, where f, g are C^∞ functions on the neighborhood. Suppose, then, that $\omega = f\,dg$ and let X, Y be C^∞-vector fields. Then, evaluating both sides of the equation of the lemma separately, we obtain

$$d\omega(X, Y) = df \wedge dg(X, Y) = df(X)\,dg(Y) - dg(X)\,df(Y)$$
$$= (Xf)(Yg) - (Xg)(Yf)$$

and

$$X\omega(Y) - Y\omega(X) - \omega([X, Y]) = X(f\,dg(Y)) - Y(f\,dg(X)) - f\,dg([X, Y])$$
$$= X(f(Yg)) - Y(f(Xg)) - f(XYg - YXg)$$
$$= (Xf)(Yg) - (Xg)(Yf)$$

after cancellation. This proves the lemma. (See Exercise 3 for a generalization.) ∎

We can state Theorem 8.3 in a more elegant way if we introduce the concept of an ideal of $\bigwedge(M)$.

(8.5) Definition An *ideal* of $\bigwedge(M)$ is a subspace $\mathscr{I}$ which has the property that whenever $\varphi \in \mathscr{I}$ and $\theta \in \bigwedge(M)$, then $\theta \wedge \varphi \in \mathscr{I}$.

(8.6) Example Let $\mathscr{J}$ be a subspace of $\bigwedge^1(M)$, that is, a collection of one-forms closed under addition and multiplication by real numbers. Then the set $\bigwedge(M) \wedge \mathscr{J} = \{\theta \wedge \varphi \mid \varphi \in \mathscr{J}\}$ is an ideal, the ideal generated by $\mathscr{J}$.

Now suppose Δ is a distribution on M and suppose that $\mathscr{I}$ is the collection of 1-forms φ on M which vanish on Δ, that is, for each $p \in M$, $\varphi_p(X_p) = 0$ for all $X_p \in \Delta_p$. $\mathscr{I}$ is a subspace; in fact, if $f \in C^\infty(M)$ and $\varphi \in \mathscr{I}$, then $f\varphi \in \mathscr{I}$. We have:

(8.7) *Δ is in involution if and only if $d\mathscr{I} = \{d\varphi \mid \varphi \in \mathscr{I}\}$ is in the ideal generated by $\mathscr{I}$.*

The proof is left to the exercises.

Exercises

These exercises involve differential forms on a manifold M. A differential form φ on M is *closed* if $d\varphi = 0$ and *exact* if $\varphi = d\theta$ for some form θ on M.

1. Show that the closed forms are a subalgebra (over $\boldsymbol{R}$) of $\bigwedge(M)$, which contains the collection of exact forms as an ideal. If $F: M \to N$ is C^∞, then show that closed forms are mapped to closed forms and exact forms to exact forms by F^*.
2. Let $M = \boldsymbol{R}^3$ and determine which of the following are closed and which are exact:
   (a) $\varphi = yz\,dx + xz\,dy + xy\,dz$;
   (b) $\varphi = x\,dx + x^2y^2\,dy + yz\,dz$;
   (c) $\theta = 2xy^2\,dx \wedge dy + z\,dy \wedge dz$.
3. Show that the following generalization of Lemma 8.4 is true for every $\varphi \in \bigwedge^r(M)$:

$$d\varphi(X_1, \ldots, X_{r+1}) = \sum_{i=1}^{r+1} (-1)^{i-1} X_i \varphi(X_1, \ldots, \hat{X}_i, \ldots, X_{r+1})$$
$$+ \sum_{i<j} (-1)^{i+j} \varphi([X_i, X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{r+1})$$

   (where the caret means that the term is omitted).
4. Prove (8.7).

A C^∞ vector field X on a manifold M determines two operators on differential forms, i_X: $\bigwedge^r(M) \to \bigwedge^{r-1}(M)$ and L_X: $\bigwedge^r(M) \to \bigwedge^r(M)$, as follows (recall, $\bigwedge^0(M) = C^\infty(M)$). If $\varphi \in \bigwedge^r(M)$, then

> $(i_X\varphi)(X_1, \ldots, X_{r-1}) = \varphi(X, X_1, \ldots, X_{r-1})$ if $r \geq 1$, and $i_X = 0$ on $\bigwedge^0(M)$; $L_X\varphi = i_X\,d\varphi + d(i_X\varphi)$ if $r \geq 1$, and $L_X f = Xf$ for $f \in \bigwedge^0(M)$.

The following exercises, some of which are difficult, develop some essential properties of these operators and are excellent exercises in the calculus of differential forms (see Notes below for references).

5. Show that i_X has the following properties:
   (i) $i_X: \bigwedge(M) \to \bigwedge(M)$ is $C^\infty(M)$-linear (fixed X), and $X \to i_X\varphi$ is $C^\infty(M)$-linear for fixed φ.
   (ii) $i_X^2 = 0$, and if $\varphi \in \bigwedge^r(M)$, $\psi \in \bigwedge^s(M)$, then

$$i_X(\varphi \wedge \psi) = (i_X\varphi) \wedge \psi + (-1)^r\varphi \wedge (i_X\psi).$$

6. Show that L_X has the following properties:
   (i) $L_X: \bigwedge(M) \to \bigwedge(M)$ is $\mathbf{R}$-linear (but not $C^\infty(M)$-linear).
   (ii) $L_X d = dL_X$.
   (iii) If $\varphi \in \bigwedge^r(M)$ and $\psi \in \bigwedge^s(M)$, then

$$L_X(\varphi \wedge \psi) = (L_X\varphi) \wedge \psi + \varphi \bigwedge(L_X\psi).$$

   (iv) $L_{fX}\varphi = df \wedge i_X\varphi + fL_X\varphi$ for $f \in C^\infty(M)$.
7. Let $\theta: W \to M$ denote local one-parameter group action generated by X and $\varphi \in \bigwedge^r(M)$, $r \geq 0$, define $L_X\varphi$

$$(L_X\varphi)_p = \lim \frac{1}{t}[\theta_t^*\varphi_{\theta(t,p)} - \varphi_p].$$

   Show that the two definitions of $L_X\varphi$ agree.
8. If $\varphi \in \bigwedge^r(M)$ and $X_1, \ldots, X_r \in \mathfrak{X}(M)$, then verify the following formula for $L_X\varphi$:

$$(L_X\varphi)(X_1, \ldots, X_r) = X(\varphi(X_1, \ldots, X_r)) - \sum_{i=1}^{r} \varphi(X_1, \ldots, L_X X_i, \ldots, X_r).$$

Notes

A translation of Riemann's famous Inaugural Address, given at Göttingen on the occasion of his being named a Privat dozent (an instructor whose fees depended on how many students came to his lectures) may be found in the book of Spivak [2], which, as mentioned earlier, is particularly valuable source material for those interested in further reading on Riemannian geometry. Volume II of this book is built around an explanation of the often obscure meaning of Riemann's lecture. This material is also relevant to Chapters VII and VIII below.

The author does not know when partitions of unity were first introduced. They are somewhat out of place in this chapter, but were placed here since this is the first point at which they were needed. Not all of the applications, as is seen, have to do with tensors. The imbedding theorem given here is a very weak version of the Whitney imbedding theorem for which proofs of stronger versions can be found in several of the references, for example, Auslander and Mackenzie [1], Hirsch [1], Guillemin and Pollack [1], and Sternberg [1]. These proofs are within the reach of the reader at this point (only Exercise 10 of VI.1 is still needed) and would form a valuable supplement to the text, especially for those interested in differential topology. The same is true of various approximation theorems (especially those of Munkres [1, Sections 3 and 4], which will be very useful for readers who wish to pursue further the consequences of differentiable structure alone (without further geometric structure such as a Riemannian metric, Lie group structure, and so forth). This is basic to modern differentiable topology (see Milnor [2] for example and Hirsch [1] or Brocker and Janich [1]).

Readers who desire a more complete and general approach to tensors and tensor fields will find it in many of the texts listed in the references. Both Kobayashi and Nomizu [1] and Sternberg [1] begin with this subject and could be studied with profit at this point, as could any book on multilinear algebra, e.g. Bourbaki.

Exterior differential forms were first used extensively by Elie Cartan whose work has enormously influenced all modern differential geometry and Lie group theory. The calculus of $\bigwedge(M)$, the exterior algebra on M, is his creation and he made many applications of it, too numerous to discuss here. Some idea of his contributions may be found in the memorial article by Chern and Chevalley [1]. Abraham and Marsden [1] and Godbillon [1] develop the calculus of exterior forms much further than we do here in order to make applications to dynamical systems and Hamiltonian and Lagrangian mechanics. The latter has an especially elegant and concise treatment of the operators i_X, L_X, and d and their interrelations. Both books are recommended for readers who wish to master this calculus—which is essential to a deeper study of the subject.

VI INTEGRATION ON MANIFOLDS

The chapter begins with a brief review (without proofs) of properties of multiple integrals over domains of $\boldsymbol{R}^n$. In the next section this theory is extended to C^∞ manifolds. The extension to manifolds involves two steps: first, we define integrals over the entire manifold M of suitable exterior n-forms and second, for those M which have a predetermined volume element (for example, Riemannian manifolds), integrals of functions over domains are defined. All the standard properties of integrals follow readily from the corresponding facts in the Euclidean case. As an illustration of the use of integration on manifolds an application is made to compact Lie groups. It is shown that by averaging a left-invariant Riemannian metric on a compact group one may obtain a bi-invariant Riemannian metric. With the same techniques—due to Weyl—it is shown that any representation of a compact group as a matrix group acting on a vector space leaves invariant some inner product on that vector space, from which it follows that any invariant subspace has a complementary invariant subspace.

Following this, in Section 4, the concept of manifold with boundary is introduced. This generalizes the line interval, unit disk, and similar simple manifold-like objects needed if one is to discuss "pasting" together of manifolds—as in Chapter I—or differentiable homotopy. However, our immediate interest is in a statement and proof of Stokes's theorem, using manifolds with boundary as domains of integration. This theorem, a generalization of the fundamental theorem of calculus, embodies Green's theorem on the plane, the divergence theorem, and Stokes's theorem of advanced calculus in a unified form. If M is a manifold with boundary ∂M and ω an $n-1$ form on M, $\dim M = n$, then the theorem asserts the equality of the integral of ω over ∂M (with suitable orienttation) and $d\omega$ over M. This theorem, proved in Section 5, concludes our development of the basic techniques of integration on manifolds.

The remainder of the chapter is devoted to various applications of the techniques accumulated thus far to the (algebraic) topology of manifolds. In order to introduce these ideas the concept of homotopy or deformation of mappings is introduced. The simplest case is the deformation over a manifold M of a loop based at $b \in M$, that is, a continuous image of $0 \leq t \leq 1$ with both endpoints at b. In general, not all loops can be deformed to one another on M (consider $M = T^2$, for example). The classes of those which can be deformed to one another, with a suitable product, form a group—the Poincaré fundamental group of M. Although quite diverse in general, these groups are isomorphic for two homeomorphic manifolds, furnishing the simplest example of an algebraic object which measures topological invariants of a space.

Following this, the de Rham groups are defined. They are the groups of closed k-forms modulo exact k-forms, and are used here, together with integration theory, to prove some classical theorems of topology (in the spirit of Milnor [2]). In particular, a proof is given of the Brouwer fixed point theorem and of the nonexistence of nowhere vanishing vector fields on even-dimensional spheres. Finally these techniques are once more applied to compact Lie groups to obtain—by way of example only—a few interesting scraps of information about their topology.

1 Integration in $\boldsymbol{R}^n$ Domains of Integration

As might be expected, we begin with integration in Euclidean space and carry over to manifolds the basic ideas developed there, just as we have done for differential calculus in earlier chapters. The basic facts that we will need concerning integrals on various subsets of $\boldsymbol{R}^n$ will be assumed known. We shall enumerate them here, and they may be found, proved in detail, in the references, for example, Apostol [1] or Spivak [1]. We need only the Riemann integral. However, we must admit domains of integration and functions which are slightly more complicated than those found in elementary calculus. This is natural since a diffeomorphism, or change of coordinates, badly distorts even a simple region such as a cube. First, we proceed to define the domains of integration which we allow.

We shall say that a subset A of $\boldsymbol{R}^n$ has (n-dimensional) *Jordan content zero*, $c(A) = 0$, if for any $\varepsilon > 0$, there exists a *finite* collection of cubes $C_1, \ldots, C_s$ which cover A and the sum of whose volumes is less than ε, $\sum_{i=1}^{s} \operatorname{vol} C_i < \varepsilon$. If A satisfies a similar condition with the less rigid requirement that for $\varepsilon > 0$ there exists a *countable* set of cubes covering A with $\sum_{i=1}^{\infty} \operatorname{vol} C_i < \varepsilon$, then we say that A has *Lebesgue measure zero*, $m(A) = 0$. These are not equivalent concepts. It is easy to see, for example, that the subset of rational numbers in $\boldsymbol{R}$ has measure zero but not content zero. However, $c(A) = 0$ implies $m(A) = 0$ and, if A is compact, the converse also holds. More generally $m(A) = 0$ if and only if A is a countable union of sets of content zero.

(1.1) Definition A bounded subset D of $\boldsymbol{R}^n$ is said to be a *domain of integration* if its boundary Bd D has content zero. A function f on $\boldsymbol{R}^n$ is said to be *almost continuous* if the set of points at which it fails to be continuous has content zero.

The most obvious example of a domain of integration is a cube, or an n-ball. The usual domains of integration in $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$, bounded by piecewise differentiable curves or surfaces, are also examples.

(1.2) Theorem *Let D be a domain of integration in $\boldsymbol{R}^n$ and let f be a real-valued function on D. Suppose that f is bounded and almost continuous on D. Then the Riemann integral $\int_D f\,dv$ exists.*

We shall refer to a function with these properties as *integrable on D*. To say that the integral exists means, of course, that it is a limit of approximating sums in the usual sense. The proof is essentially the same as that which is at least outlined in every calculus book. It is a very useful exercise to carry it out in detail and then to verify the following properties which are relatively easy consequences of the reasoning used in proving existence.

Basic Properties of the Riemann Integral

Let D, D_1, and D_2 denote domains of integration in $\boldsymbol{R}^n$ and f, g bounded almost continuous functions on $\boldsymbol{R}^n$. It is not too difficult to show that $\bar{D}$, the closure of D, and $\mathring{D}$, the interior of D, are also domains of integration as are $D_1 \cup D_2$, $D_1 \cap D_2$, and $D_1 - D_2$. We have the following standard properties:

(1.3) *If $c(D) = 0$, then* $\displaystyle\int_D f\,dv = 0.$

(1.4) $$\int_{D_1 \cup D_2} f\,dv = \int_{D_1} f\,dv + \int_{D_2} f\,dv - \int_{D_1 \cap D_2} f\,dv.$$

(1.5) $$\int_D (af + bg)\,dv = a\int f\,dv + b\int g\,dv \qquad \textit{for all} \quad a, b \in \boldsymbol{R}.$$

(1.6) *Suppose $f \geq 0$ on D and $c(D) \neq 0$. Then $\int_D f\,dv \geq 0$. Equality holds if and only if $f = 0$ at every point at which it is continuous.*

Recall that the *characteristic function* k_A of a subset A of a space X is defined to be identically equal to $+1$ on A and 0 outside A, that is, on the complement of A. Therefore k_A is bounded and its discontinuities are exactly the set of boundary points of A, Bd A. In particular, if D is a domain of integration, we have $c(\text{Bd } D) = 0$ so that k_D is integrable. If D' is a domain of integration which contains D, then $\int_{D'} k_D f\,dv = \int_D f\,dv$. Thus if f on $\boldsymbol{R}^n$ is

bounded, has compact support, and is almost continuous, then we define $\int_{R^n} f\,dv$ unambiguously by $\int_{R^n} f\,dv = \int_D f\,dv$, using any domain of integration D such that $D \supset \operatorname{supp} f$.

(1.7) Definition Let D be any domain of integration. Then we define the *volume of* D, vol D, by

$$\operatorname{vol} D = \int_{R^n} k_D\,dv = \int_D k_D\,dv.$$

The following property is an easy consequence of the definitions:

(1.8) $$\left(\inf_D f\right) \operatorname{vol} D \le \int_D f\,dv \le \left(\sup_D f\right) \operatorname{vol} D.$$

When D is connected and f is continuous, we obtain the mean value property

$$\int_D f\,dv = f(a) \operatorname{vol} D$$

for some point $a \in D$.

The following theorem, a special case of Fubini's theorem, is more difficult to prove than the above properties, although we need only the simplified version below. It justifies the usual evaluation of multiple integrals by repeated single integrations of functions of one variable (iterated integrals).

(1.9) Theorem *If f is a continuous function on the domain of integration $D = \{x \in R^n\colon a^i \le x^i \le b^i,\ i = 1, \ldots, n\}$, then*

$$\int_D f\,dv = \int_{a^n}^{b^n} \cdots \int_{a^1}^{b^1} f(x^1, \ldots, x^n)\,dx^1 \cdots dx^n,$$

the expression on the right denoting repeated single integrations.

We shall need one further theorem from advanced calculus, the principle which allows us to change the variable, or variables, of integration. In the case of a function of one variable this is the standard and indispensable technique of *substitution*, which allows us to write

$$\int_c^d f(y)\,dy = \int_a^b f(g(x)) \frac{dy}{dx}\,dx,$$

where $y = g(x)$, $a \le x \le b$, with $c = g(a)$ and $d = g(b)$. Unless we assume $dy/dx > 0$, we encounter problems in this formula. If this condition is

satisfied, then $y = g(x)$ may be considered as a change of variable, or as a diffeomorphism of $[a, b]$ onto $[c, d]$. The general multiple-integral statement is less familiar in elementary calculus, although it occurs, for example, in the passage from Cartesian to polar, cylindrical, or spherical coordinates. It is proved in most advanced calculus courses for $n = 2$ or 3 at least, so we will assume it without proof. This is the most difficult of the standard theorems of integral calculus which we will expect of the reader. It will be essential to us in extending Riemann integration to manifolds, since we clearly must know the effect of change of coordinates on the value of an integral.

Let us denote by $G: U \to U'$ a diffeomorphism of $U \subset \boldsymbol{R}^n$ onto $U' \subset \boldsymbol{R}^n$ and by ΔG the determinant of its Jacobian. We suppose G to be given by coordinate functions $y^i = y^i(x)$, $i = 1, \ldots, n$. Then $\Delta G = \det(\partial y^i/\partial x^j)$. A function f' on U' determines a function $f = f' \circ G$ on U and we have the following relation of their integrals.

(1.10) Theorem (Change of Variables) *Suppose* $D \subset U$ *and* $D' = G(D) \subset U'$ *are domains of integration and that* f' *is integrable on* D'. *Let* $f = f' \circ G$, *that is,* $f(x^1, \ldots, x^n) = f'(g^1(x), \ldots, g^n(x))$. *Then* f *is integrable on* D *and*

$$\int_{D'} f'(y)\, dv' = \int_D f'(G(x))\,|\Delta G|\; dv = \int_D f(x)\,|\Delta G|\; dv.$$

(1.11) Example Let

$$D = \{\rho, \theta, \varphi)|0 < a \le \rho \le b, 0 \le \theta \le \pi/2, \pi/4 \le \varphi \le \pi/2\}$$

and D' be the first quadrant region of xyz-space between the spheres with center at the origin and radii a and b, and outside the inverted cone $z^2 = x^2 + y^2$ (Fig. VI.1). Let G be given by the coordinate functions

$$x = \rho \sin \varphi \cos \theta, \qquad y = \rho \sin \varphi \sin \theta, \qquad z = \rho \cos \varphi.$$

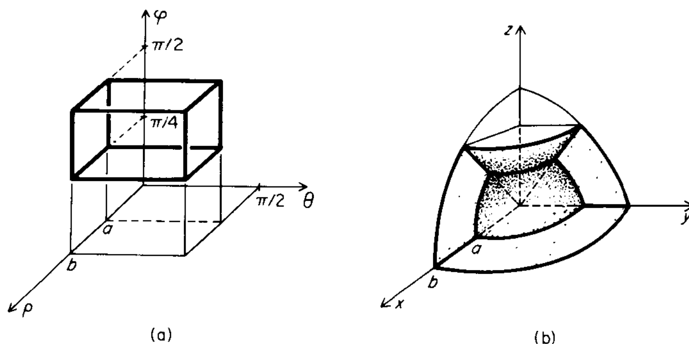


Figure VI.1

Given $f'(x, y, z) = x^2 + y^2 + z^2$, then $f = f' \circ G$ is

$$f(\rho, \theta, \varphi) = f'(\rho \sin \varphi \cos \theta, \rho \sin \varphi \sin \theta, \rho \cos \varphi) = \rho^2$$

and $\Delta = \rho^2 \sin \varphi$ so that

$$\int_{D'} (x^2 + y^2 + z^2)\, dx\, dy\, dz = \int_D \rho^2 \,|\rho^2 \sin \varphi|\; d\rho\, d\varphi\, d\theta.$$

If we are to extend these ideas to C^∞ manifolds, we need to know what happens to domains of integration under diffeomorphisms. A cube in $\boldsymbol{R}^n$, for example, is such a domain since its boundary—the faces—have zero (n-dimensional) content. Does it remain a domain of integration after a diffeomorphism? If we recall that there are continuous images of an interval, $0 \le t \le 1$, which fill a square, this question seems less trivial; it is possible that the image of the boundary of a cube could become very large under a differentiable mapping. The following lemma will thus be crucial

(1.12) Lemma *Let A be a compact subset of $\boldsymbol{R}^n$ and let $F\colon A \to \boldsymbol{R}^m$, $n \le m$ be a C^1 mapping. Then $F(A)$ has content zero if $n < m$ or if $n = m$ and A has content zero.*

Proof By definition F must be C^1 on an open set $U \supset A$ and since A is compact, we may choose an open set $V \supset A$ such that $\bar{V}$ is a compact subset of U. Let $K = \sup_{x \in \bar{V}} |\partial f^i / \partial x^j|$, a bound of the derivatives on $\bar{V}$ of the coordinate functions of the map F. Choose δ_1, $0 < \delta_1 \le 1$, so that every cube of side δ_1 whose center is in A lies inside V. By the mean value theorem (Theorem II.2.2), we have

$$\|F(x) - F(a)\| < (nm)^{1/2} K \|x - a\|$$

for any x in a cube of side δ_1 and center $a \in A$. If $\delta_1 > \delta > 0$, then a cube C of side δ and center $a \in A$ must map into a cube C' of center $F(a)$ and side length less than or equal to $(nm)^{1/2} K\delta$. Thus we see that $F(C)$ lies in a cube C' whose volume satisfies

$$\operatorname{vol} C' \le ((nm)^{1/2} K\delta)^m = (nm)^{m/2} K^m \delta^{m-n} \delta^n \le k\delta^{m-n} \operatorname{vol} C,$$

where $k = K^m (nm)^{m/2}$ is independent of $a \in A$ and $\operatorname{vol} C = \delta^n$. Assume $m = n$ and $c(A) = 0$. From this it follows at once that given any $\varepsilon > 0$, we may cover $F(A)$ with a finite number of cubes $C_1', \ldots, C_s'$ whose total volume is less than ε. We need only cover A with cubes $C_1, \ldots, C_s$ whose total volume is less than ε/k and whose side is less than δ_1. This shows that the content of $F(A)$ is zero in this case. The case $n < m$ is left to the reader. For an extension see Exercise 10. ∎

Using this lemma it is easy to extend the notions of zero content and zero measure to subsets of any C^∞ manifold M of dimension n.

(1.13) Definition A subset $A \subset M$ is said to have *content zero*, $c(A) = 0$, if it is contained in the union of a finite number of compact subsets A_i, $A \subset A_1 \cup \cdots \cup A_s$, each of which lies in a coordinate neighborhood U_i, φ_i such that $c(\varphi_i(A_i)) = 0$ in $\boldsymbol{R}^n$, $i = 1, \ldots, s$. An arbitrary subset $B \subset M$ is said to have *measure zero*, $m(B) = 0$, if B is the union of a countable collection of subsets $B = \bigcup_{i=1}^{\infty} B_i$ such that each B_i has content zero.

In the light of this definition we may state, as a corollary to the lemma, the following facts about sets on a manifold:

(1.14) Corollary *If $A \subset M$ has content (measure) zero and $F: M \to N$ is a C^1 map with* $\dim M \leq \dim N$, *then $F(A)$ has content (measure) zero. In particular, this holds if F is a diffeomorphism.*

Proof This is an obvious application of Lemma 1.12 to Definition 1.13. If $\dim M < \dim N$, a stronger result holds (Exercise 10). ∎

We define domain of integration in an arbitrary manifold precisely as we did for $\boldsymbol{R}^n$: $D \subset M$ is a *domain of integration* if D is relatively compact and the boundary of D has content zero, $c(\text{Bd } D) = 0$. [Note that in $\boldsymbol{R}^n$ "relatively compact" is equivalent to "bounded."] We have analogous properties to those of domains of integration in $\boldsymbol{R}^n$.

(1.15) Theorem *If D is a domain of integration in M, so are its closure and its interior. Finite unions and intersections of domains of integration are domains of integration and the image of a domain of integration under a diffeomorphism is a domain of integration.*

Proof These are all immediate consequences of Definition 1.13 and of the corresponding statements for subsets of content zero and domains of integration in $\boldsymbol{R}^n$. For the last statement we must note that a diffeomorphism takes boundary points to boundary points. ∎

Exercises

1. Prove that a set of measure zero cannot contain any open set.
2. Prove Theorem 1.2 for a general domain of integration in $\boldsymbol{R}^2$ assuming that the integrand f is: (a) continuous on D, and then (b) continuous except for a set of content zero.
3. Prove that finite unions and intersections of domains of integration are domains of integration.
4. If D is a domain of integration show that $\bar{D}$, its closure, and $\mathring{D}$, its interior, are domains of integration and that if f is integrable on D, then

$$\int_D f\,dv = \int_{\bar{D}} f\,dv = \int_{\mathring{D}} f\,dv.$$

5. Verify properties (1.3)–(1.6) of the Riemann integral.

6. Let D be a domain of integration in $\boldsymbol{R}^n$ and $a < t < b$ an open interval of $\boldsymbol{R}$. Suppose $f(x, t)$ is continuous on $\bar{D} \times (a, b)$ and is of class C^1 in t. Then prove that $g(t) = \int_D f(x, t)\, dv$ is of class C^1 on (a, b) and $dg/dt = \int_D (\partial f/\partial t)\, dv$.
7. Suppose f, g are both integrable on a domain of integration D and that $f \geq g$. Then show that $\int_D f\, dv \geq \int_D g\, dv$.
8. Prove the change of variables theorem for a linear mapping $G\colon \boldsymbol{R}^2 \to \boldsymbol{R}^2$.
9. If $\{f_n(x)\}$ is a sequence of continuous functions on D (a domain of integration in $\boldsymbol{R}^n$) converging uniformly to $g(x)$, show that $\lim_{n\to\infty} \int_D f_n(x)\, dv = \int_D g(x)\, dv$.
10. Prove that if $F\colon M \to N$ is a C^1 mapping and $\dim N > \dim M$, then $F(M) \subset N$ has measure zero.

2 A Generalization to Manifolds

In this section we carry over to arbitrary oriented manifolds the concept of integral reviewed in the previous section. We first define the integral of an n-form ω on the oriented manifold M of dimension n, denoting it by $\int_M \omega$. It is only when we specialize to a more restricted class of manifolds, say Riemannian manifolds, that we are able to give meaning to the integral of a *function* on M over a domain D of integration in M and thus obtain a complete generalization of integrals on $\boldsymbol{R}^n$. This is not surprising since definition of the Riemann integral in $\boldsymbol{R}^n$ makes important use of *volume*, a metric concept, which is not defined on a general differentiable manifold.

Suppose that M is an *oriented* manifold and $\dim M = n$. By Definition V.7.5 this means that there is a C^∞ n-form Ω on M which is not zero at any point of M. It is called a *volume element*; any other n-form ω is given by $\omega = f\Omega$, where f is a function on M. Since Ω is C^∞, ω will have the differentiability class of f. We use this to make the following definitions.

(2.1) Definition A function f on M is *integrable* if it is bounded, has compact support (vanishes outside a compact set), and is almost continuous (that is, continuous except possibly on a set of content zero). An n-form ω on M, in the very general sense of a function assigning to each $p \in M$ an element ω_p of $\bigwedge^n(T_p(M))$, is said to be *integrable* if $\omega = f\Omega$, where f is an integrable function. [Note: We are not requiring ω to be C^∞ or even C^1.]

We remark that the definition of integrable n-form does not depend on the particular Ω we use. Any other $\tilde{\Omega}$ giving the orientation is of the form $\tilde{\Omega} = g\Omega$, where g is a *positive* C^∞ function on M; thus $f\Omega = f/g\tilde{\Omega}$. If f has compact support, is bounded, and is almost continuous, then the same will be true of f/g. We will denote by $\bigwedge_0^n(M)$ the set of integrable n-forms. Like

$\bigwedge^n(M)$, it is a vector space over $\boldsymbol{R}$; moreover, it is closed under multiplication by continuous or integrable functions on M.

We shall refer to a subset $Q \subset M$ as a *cube* of M if it lies in the domain of an associated, *oriented*, coordinate neighborhood U, φ and $\varphi(Q) = C = \{x \in \boldsymbol{R}^n \mid 0 \le x^i \le 1, i = 1, \ldots, n\}$, the unit cube of $\boldsymbol{R}^n$. Thus a cube is a compact set and is coordinatized in a definite way. We first define the integral over M of any $\omega \in \bigwedge_0^n(M)$ whose support lies interior to some cube Q. Let U, φ be the coordinate neighborhood associated with Q and suppose

$$\varphi^{-1*}(\omega) = f(x)\, dx^1 \wedge \cdots \wedge dx^n$$

represents ω in the local coordinates. Then f is bounded and almost continuous on C so that $\int_C f\, dv$ is defined. We define

$$\int_M \omega = \int_C f\, dv.$$

We must show that the value of this integral is independent of the particular cube we have used. Suppose Q' is another cube containing supp ω and let U', φ' be the associated coordinate neighborhood. We denote the local coordinates for this neighborhood by $y^1, \ldots, y^n$ and suppose that

$$\varphi'^{-1*}(\omega) = f'(y)\, dy^1 \wedge \cdots \wedge dy^n$$

represents ω on $\varphi'(U')$. According to the rules for change of components of an n-form, we have

$$f(x) = f'(G(x))\, \Delta G,$$

where $G = \varphi' \circ \varphi^{-1}: \varphi(U \cap U') \to \varphi'(U \cap U')$ and ΔG is the determinant of the Jacobian matrix of this diffeomorphism; ΔG is positive since these are oriented neighborhoods (see Section V.7). On the other hand, since Q, Q' are domains of integration, so are $Q \cap Q'$ and its images $D = \varphi(Q \cap Q')$ and $D' = \varphi'(Q \cap Q')$ which lie in the unit cube of the x-coordinate space and the y-coordinate space, respectively. Moreover, $\operatorname{supp} \omega \subset Q \cap Q'$, so $\operatorname{supp} f \subset D$ and $\operatorname{supp} f' \subset D'$. Therefore

$$\int_C f(x)\, dv = \int_D f(x)\, dv \qquad \text{and} \qquad \int_{C'} f'(y)\, dv' = \int_{D'} f'(y)\, dv'.$$

According to the change of variable theorem 1.10, and since $D' = G(D)$, we have

$$\int_{D'} f'(y)\, dv' = \int_D f'(G(x))\, |\Delta G|\, dv.$$

However, $\Delta G > 0$ so that $|\Delta G| = \Delta G$ and the integral on the right must then be equal to $\int_D f(x)\, dv$ by the above formula for change of components. This shows that $\int_M \omega$ is uniquely determined for every integrable ω which

vanishes outside of some cube. We note, in particular, the following linearity property: If ω_1, ω_2 vanish outside a cube Q and a_1, a_2 are real numbers, then

$$\int_M a_1\omega_1 + a_2\omega_2 = a_1 \int_M \omega_1 + a_2 \int_M \omega_2 .$$

Now suppose that ω is an arbitrary integrable n-form. We will define $\int_M \omega$ in this case by representing ω as the sum of a finite number of forms of the special type above and adding their integrals to obtain $\int_M \omega$. More precisely, let $K = \operatorname{supp} \omega$ and choose a finite covering of K by the interiors $\mathring{Q}_1, \ldots, \mathring{Q}_b$ of cubes $Q_1, \ldots, Q_b$ associated with coordinate neighborhoods $U_1, \varphi_1, \ldots, U_b, \varphi_b$, respectively. The open sets $M - K, \mathring{Q}_1, \ldots, \mathring{Q}_b$ cover M. Take a partition of unity $\{f_i\}$ subordinate to this covering. By compactness of K and the local finiteness of $\{\operatorname{supp} f_i\}$, there is an s such that for $i > s$, $K \cap \operatorname{supp} f_i = \varnothing$, so that $f_i\,\omega = 0$. Since $\sum f_i \equiv 1$, we have then

$$\omega = f_1\omega + \cdots + f_s\omega$$

and we define

$$\int_M \omega = \int_M f_1\omega + \cdots + \int_M f_s\omega.$$

Each of the integrals on the right is defined since the integrand has its support on the interior $\mathring{Q}_j$ of a cube Q_j.

The value of this integral does not depend on the choice of covering or the functions $\{f_i\}$. Let $Q'_1, \ldots, Q'_{b'}$ be another set of cubes whose interiors cover K and choose a partition of unity $\{g_k\}$ subordinate to $M - K$, $\mathring{Q}'_1, \ldots, \mathring{Q}'_{b'}$. Again, $g_k\omega \equiv 0$ if k is greater than some r. Then $\sum_{i,k} f_i g_k \equiv \sum_i f_i \sum_k g_k \equiv 1$ and for fixed k, $1 \le k \le r$, we have $\operatorname{supp} f_i g_k \subset Q'_l$ for some l. Therefore

$$\int_M g_k\omega = \int_M f_1 g_k\omega + \cdots + \int_M f_s g_k\omega$$

by the linearity of the integral with respect to forms with support in the same cube. Therefore, if we compute $\int_M \omega$ using this second covering by cubes, we have

$$\int_M \omega = \sum_{k=1}^{r} \int_M g_k\omega = \sum_{k=1}^{r} \sum_{i=1}^{s} \int_M f_i g_k\omega.$$

However, by a symmetric argument the sum on the right is also equal to $\sum_{i=1}^{s} \int_M f_i\,\omega$, hence both choices assign the same value to $\int_M \omega$. This completes the definition of the integral over M of integrable n-forms, we now list some of its properties.

(2.2) Theorem *The process just defined assigns to each integrable n-form ω on an oriented manifold M a real number $\int_M \omega$. We have the following properties:*

(i) *If $-M$ denotes the same underlying manifold with opposite orientation, then $\int_{-M} \omega = -\int_M \omega$.*

(ii) *The mapping $\omega \to \int_M \omega$ is an $\boldsymbol{R}$-linear mapping on $\bigwedge_0^n(M)$, that is:*

$$\int_M a_1\omega_1 + a_2\omega_2 = a_1 \int \omega_1 + a_2 \int \omega_2 ,$$

$$a_1 a_2 \in \boldsymbol{R} \quad \text{and} \quad \omega_1, \omega_2 \in \textstyle\bigwedge_0^n(M).$$

(iii) *If Ω is a nowhere vanishing n-form giving the orientation of M and $\omega = g\Omega$ with $g \geq 0$, then $\int_M g\Omega \geq 0$ and equality holds if and only if $g = 0$ wherever it is continuous.*

(iv) *If $F: M_1 \to M_2$ is a diffeomorphism and $\omega \in \bigwedge_0^n(M_2)$, then*

$$\int_{M_1} F^*\omega = \pm \int_{M_2} \omega,$$

with sign depending on whether F preserves or reverses orientation.

Proof Because of the definition of the integral, we need to verify these properties only for forms ω whose support lies in a cube Q associated with the oriented coordinate neighborhood U, φ and coordinates $x^1, \ldots, x^n$. Then by definition, $\int_M \omega = \int_C f(x)\, dv$, where $\varphi^{-1*}(\omega) = f(x)\, dx^1 \wedge \cdots \wedge dx^n$. If orientation of M is reserved, in evaluating $\int_M \omega$ we replace U, φ by, say, $U' = U, \varphi'$ with φ' such that the Jacobian of $\varphi' \circ \varphi^{-1}$ has negative determinant, e.g., φ followed by interchange of first and second variables. This changes the sign of f since f is the component of ω in the local coordinates, hence it changes the sign of the integral. Property (ii) was previously noted; it is a consequence of the corresponding property for the Riemann integral on $\boldsymbol{R}^n$. Property (iii) is clear once we note that in (oriented) local coordinates $\varphi^{-1*}\Omega = p(x)\ dx^1 \wedge \cdots \wedge dx^n$, $p(x) > 0$, so that $\int_M g\Omega = \int_C g(x)p(x)\ dv$. Since $g(x)p(x) \geq 0$, and vanishes exactly where $g(x)$ vanishes, the assertion follows from the corresponding property in $\boldsymbol{R}^n$. Finally, suppose $F: M_1 \to M_2$ is a diffeomorphism which preserves orientation. If ω on M_2 has support in a cube Q associated with the coordinate neighborhood U, φ, then $Q' = F^{-1}(Q)$ is a cube on M_1 associated with $U' = F^{-1}(U)$ and $\varphi' = \varphi \circ F^{-1}$. Using this cube, which contains the support of $F^*\omega$, we have precisely the same expression $f(x)\ dx^1 \wedge \cdots \wedge dx^n$ for both ω and $F^*\omega$ in local coordinates, hence the same integral $\int_C f\ dv$ gives the value of both $\int_{M_2} \omega$ and $\int_{M_1} F^*\omega$. If F does not preserve orientation, the equation $\int_{M_1} F^*\omega = -\int_{M_2} \omega$ follows from the orientation-preserving case and property (i). ∎

(2.3) Remark We note that a special case of the definition above, namely $M = \boldsymbol{R}^n$, defines

$$\int_{\boldsymbol{R}^n} f(x^1, \ldots, x^n)\, dx^1 \wedge \cdots \wedge dx^n$$

for any bounded function f on $\boldsymbol{R}^n$ which has compact support and is almost continuous. It is left as an exercise to show that if $\operatorname{supp} f \subset D$, a domain of integration, then

$$\int_{\boldsymbol{R}^n} f(x)\, dx^1 \wedge \cdots \wedge dx^n = \int_D f(x)\, dv,$$

the usual Riemann integral.

Integration on Riemannian Manifolds

Thus far we have not defined integrals of *functions*, but rather integrals of *n-forms*. Even a cursory examination of the definition of integrals of functions over domains of $\boldsymbol{R}^n$ used in advanced calculus shows that it assumes that we are able to assign a volume to certain classes of subsets of $\boldsymbol{R}^n$, say, cubes and rectangular parallelepipeds. In fact only this one ingredient is lacking; if M has a well-determined volume element, then we are able to pass from the definition above to integrals of functions on an oriented manifold M. A *volume element* is, by definition, a nowhere vanishing n-form Ω on M which is in that class which determines the orientation. On an arbitrary oriented manifold there is such a form Ω but it is determined only to within a multiple by a positive C^∞ function. This is not enough to define volumes; we must have a unique Ω given, say, by the structure of M. One case in which this occurs, according to Theorem V.7.7, is on an oriented Riemannian manifold M. In this case there is a unique Ω whose value on any orthonormal frame is $+1$. We shall always use this Ω on the Riemannian manifold and in the remainder of this section we shall discuss only the Riemannian case. Then, using Ω and the characteristic function k_D of a domain of integration D we are able to parallel the theory for $\boldsymbol{R}^n$.

(2.4) Definition If D is a domain of integration on an oriented Riemannian manifold M and k_D is the characteristic function of D, we define the *volume* of D, denoted by vol D, by vol $D = \int_M k_D \Omega$. If f is any bounded, almost continuous function on M, we define the integral of f over D, denoted $\int_D f$, by $\int_D f = \int_M f k_D \Omega$. When M is compact, we may take $D = M$ and obtain vol $M = \int_M \Omega$ and $\int_M f = \int_M f\Omega$.

These integrals are defined since k_D is continuous except on Bd D, which has content zero, and k_D has compact support.

(2.5) Lemma *With these definitions the integral of f on a domain of integration on M satisfies properties (1.3)–(1.6) of the Riemann integral on $\boldsymbol{R}^n$. It is equal to the Riemann integral when $M = \boldsymbol{R}^n$ (with its standard metric).*

This is a consequence of the definitions and of the corresponding properties (1.3)–(1.6) of the Riemann integral. One merely needs to demonstrate—by choosing a covering of D by the interiors of cubes and taking a corresponding partition of unity as in the definition of $\int_M \omega$—that it is possible to reduce the proof to verifying each property for the special case in which $\omega = f\Omega$ has its support in a single cube. In this case the properties coincide with the properties of the integral on $\boldsymbol{R}^n$. For the last statement we use Remark 2.3.

We recall that in local coordinates U, φ with coordinate frames $E_1, \ldots, E_n$ and Riemannian metric tensor $\Phi(X, Y)$, the matrix components $\Phi(E_i, E_j)$ on U are customarily denoted by g_{ij}, $i, j = 1, \ldots, n$, with the same symbols g_{ij} frequently used to denote $g_{ij}(p) = \Phi_p(E_{ip}, E_{jp})$ and $\hat{g}_{ij}(x^1, \ldots, x^n) = \hat{g}_{ij}(\varphi(p))$, that is, the components considered as functions on $U \subset M$ or as the corresponding functions on $\varphi(U) \subset \boldsymbol{R}^n$. In Section V.7 we found that the local expression for Ω on an *oriented* neighborhood was

$$\varphi^{-1*}\Omega = \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n, \qquad g = \det(g_{ij}).$$

We use this in the example below.

(2.6) Example Let M be a surface in $\boldsymbol{R}^3$ with the Riemannian metric induced by the standard metric of $\boldsymbol{R}^3$ and let U, φ be a coordinate neighborhood with coordinates (u, v). Suppose $\varphi(U) = W$ an open subset of the uv-plane. Let $F = \varphi^{-1}$ so that $F: W \to M$ has image U, and let $F(u, v) = (f(u, v), g(u, v), h(u, v))$ be the C^∞-coordinate functions for the mapping (see Fig. VI.2). As in Example IV.1.10 the coordinate frames E_1, E_2 on U are

$$E_1 = F_*\left(\frac{\partial}{\partial u}\right) = \frac{\partial f}{\partial u}\frac{\partial}{\partial x} + \frac{\partial g}{\partial u}\frac{\partial}{\partial y} + \frac{\partial h}{\partial u}\frac{\partial}{\partial z},$$

$$E_2 = F_*\left(\frac{\partial}{\partial v}\right) = \frac{\partial f}{\partial v}\frac{\partial}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial}{\partial x} + \frac{\partial h}{\partial v}\frac{\partial}{\partial z},$$

and hence

$$g_{11}(u, v) = \left(\frac{\partial f}{\partial u}\right)^2 + \left(\frac{\partial g}{\partial u}\right)^2 + \left(\frac{\partial h}{\partial u}\right)^2 = (E_1, E_1),$$

$$g_{12}(u, v) = \frac{\partial f}{\partial u}\frac{\partial f}{\partial v} + \frac{\partial g}{\partial u}\frac{\partial g}{\partial v} + \frac{\partial h}{\partial u}\frac{\partial h}{\partial v} = (E_1, E_2) = (E_2, E_1) = g_{21}(u, v),$$

$$g_{22}(u, v) = \left(\frac{\partial f}{\partial v}\right)^2 + \left(\frac{\partial g}{\partial v}\right)^2 + \left(\frac{\partial h}{\partial v}\right)^2 = (E_2, E_2).$$

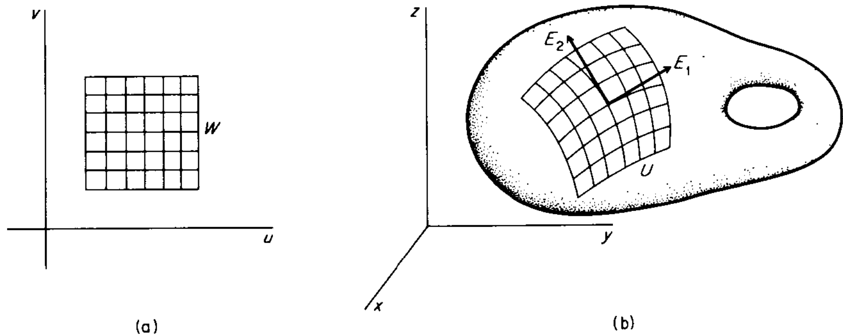


Figure VI.2

These are denoted E, F, G, respectively, in the literature of classical differential geometry and we have then

$$\varphi^{-1*}\Omega = F^*\Omega = (g_{11}g_{22} - g_{12}^2)^{1/2}\, du \wedge dv = (EG - F^2)^{1/2}\, du \wedge dv.$$

If D is a domain of integration on M such that $D \subset U$, and h is an integrable function on D, then

$$\int_D h = \int_D h\Omega = \int_{\varphi(D)} h(u, v)(EG - F^2)^{1/2}\, du \wedge dv$$

$$= \int_{\varphi(D)} h(u, v)(EG - F^2)^{1/2}\, du\, dv.$$

Suppose, for example, that φ is the (diffeomorphic) projection of an open set U of M onto an open set W of the xy-plane, which we identify with the parameter plane. In this case $F\colon W \to U$ is given by $F(x, y) = (x, y, f(x, y))$. The graph of $z = f(x, y)$ lying over W is the subset U of M, Fig. VI.3. The

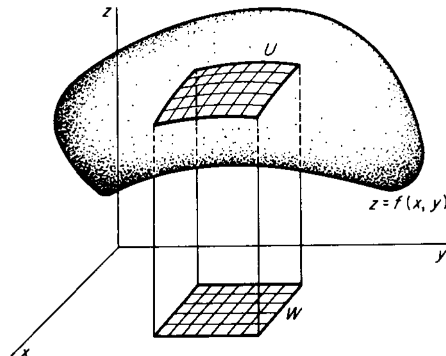


Figure VI.3

coordinate frames are $E_1 = \partial/\partial x + f_x\,\partial/\partial z$ and $E_2 = \partial/\partial y + f_y\,\partial/\partial z$, so $E = 1 + f_x^2$, $F = f_x f_y$, $G = 1 + f_y^2$. Hence

$$F^*\Omega = (EG - F^2)^{1/2}\, dx \wedge dy = (1 + f_x^2 + f_y^2)^{1/2}\, dx \wedge dy.$$

If $D \subset U$ is a domain of integration and $A \subset W$ its projection to the xy-plane, then for any integrable function h on M we have

$$\int_D h = \int_A h(x, y, z)(1 + f_x^2 + f_y^2)^{1/2}\, dx\, dy.$$

When $h \equiv 1$, the value of this integral is the area of D [$=$ vol D].

If, for example $M = S^2$, the unit sphere, let U be the upper hemisphere and $D = U$. Then $A = W = \{(x, y) \mid x^2 + y^2 < 1\}$ and $F(x, y) = (x, y, (1 - x^2 - y^2)^{1/2})$. The area of U is

$$\int_U \Omega = \int_A (1 - x^2 - y^2)^{-1/2}\, dx \wedge dy$$
$$= \int_{-1}^{+1} \int_{-(1-y^2)^{\frac{1}{2}}}^{(1-y^2)^{\frac{1}{2}}} (1 - x^2 - y^2)^{-1/2}\, dx\, dy = 2\pi.$$

(2.7) Remark In practice (or for theoretical purposes) one might hope that a compact manifold M could be covered by a finite number of domains of integration $D_1, \ldots, D_s$ with the properties: (i) $c(D_i \cap D_j) = 0$, $i \neq j$, $i, j = 1, \ldots, s$, and (ii) each D_i lies in a coordinate neighborhood U_i, φ_i. Then, using the fact that

$$\int_M f = \int_{D_1} f + \cdots + \int_{D_s} f,$$

it would be possible to evaluate each integral on the right separately as an integral on $\varphi_i(D_i) \subset \mathbf{R}^n$,

$$\int_{D_i} f = \int_{\varphi_i(D_i)} f(x)\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n = \int_{\varphi_i(D_i)} f(x)\sqrt{g}\, dv,$$

where $f(x)$ denotes the expression for f in local coordinates and $g = \det(g_{ij})$ as in the remark preceding Example 2.6.

In fact, it can be shown that any differentiable manifold M (compact or not) can be covered with a collection of domains of integration $D_1, D_2, \ldots,$ each the diffeomorphic image of a simplex (for $n = 2$ a triangle, for $n = 3$ a tetrahedron, and so on). Moreover these domains intersect in sets of content zero. [This is part of a theorem which asserts that any C^∞ manifold is *triangulable*. An example is illustrated in Fig. VI.4.] When M is compact the number of D_i is finite. This is not a complete description of a triangulation; for more details see Singer and Thorpe [1]. However, it shows that for both practical and theoretical purposes a technique of evaluation of $\int_M f$ or $\int_M \Omega$ is available.

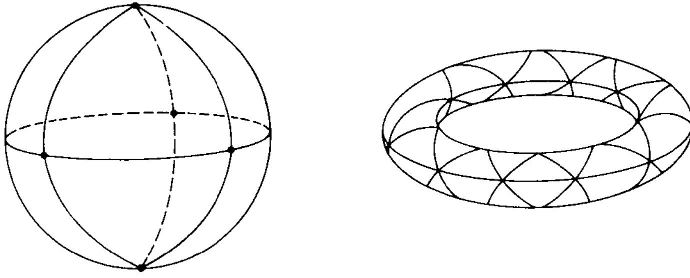


Figure VI.4
Triangulated manifolds.

Exercises

1. Interpret Ω and volume for a curve in $\boldsymbol{R}^3$, that is, a one-dimensional manifold.
2. Show that if D is a domain of integration on a manifold, then

$$\int_D f = \int_{\bar{D}} f = \int_{\mathring{D}} f.$$

3. Let D be a domain of integration in R^n and $F: \bar{D} \to M$ be a continuous mapping into an n-manifold M. Suppose F is a C^1 diffeomorphism on $\mathring{D}$ the interior of D, piecewise C^1 on ∂D, and that ω is an integrable n-form on M. Then show that $F(D)$ is a domain of integration and that $\int_{F(D)} \omega = \int_D F^*\omega$. [We do not require F to be one-to-one on the boundary of D.]
4. Find Ω (for the induced Riemannian metric) on the torus T^2 in $\boldsymbol{R}^3$ obtained by rotating a circle of radius a and center at $(b, 0, 0)$, $b > a > 0$, around the x^3-axis. Use this to determine $\mathrm{vol}(T^2)$. (Use Exercise 3.)
5. Using Remark 2.7 integrate on $M = S^2$, the unit sphere of $\boldsymbol{R}^3$, the function f giving the distance of a point on M from the plane $x^3 = -1$. Argue that we may use as D_1 and D_2 the upper and lower hemispheres. [*Hint:* Use Exercise 2.]

3 Integration on Lie Groups

One striking illustration of the uses to which integration on manifolds can be put arises when the manifold considered is a Lie group G. Although the most interesting case is a compact Lie group, for the present we allow G to be an arbitrary Lie group of dimension n. We shall need some simple observations concerning left and right translations and inner automor-

phisms of G. Given $a, b \in G$, we denote by L_a, R_b, and $I_a = L_a \circ R_{a^{-1}}$ left and right translation and the inner automorphism, $I_a(x) = axa^{-1}$, of G, respectively. These are C^∞ mappings with inverses $L_a^{-1} = L_{a^{-1}}$, $R_a^{-1} = R_{a^{-1}}$, and $I_a^{-1} = I_{a^{-1}}$. Hence they are diffeomorphisms and as such induce $\boldsymbol{R}$-linear mappings of $\mathfrak{X}(G)$—the C^∞-vector fields on G—onto itself which preserve the bracket operation (Corollary IV.7.10). However, on G our main interest is in the subspace $\mathfrak{g}$ of $\mathfrak{X}(G)$ consisting of all left-invariant vector fields on G. As we have seen $\mathfrak{g}$ is a Lie algebra, the Lie algebra of G, with respect to the product $[X, Y]$. Given $a, b \in G$, we note the fact that the left and right translations L_a and R_b commute—this is just the associative law $a(xb) = (ax)b$. From this we deduce that if $X \in \mathfrak{g}$, then $R_{b*} X \in \mathfrak{g}$. Also

$$L_{g*}(R_{b*} X) = R_{b*}(L_{g*} X) = R_{b*} X.$$

Similarly, $I_{a*} X = L_{a*} R_{a^{-1}*} X = R_{a^{-1}*} X \in \mathfrak{g}$; thus $I_{a*}: \mathfrak{g} \to \mathfrak{g}$. Because I_{a*} is both a linear mapping and preserves the product, that is, $I_{a*}[X, Y] = [I_{a*} X, I_{a*} Y]$, it is an automorphism of the Lie algebra $\mathfrak{g}$. Finally, note that $I_{ab} = I_a \circ I_b$ so that $I_{ab*} = I_{a*} \circ I_{b*}$ by the chain rule. Putting these facts together and adopting the notation $\operatorname{Ad} g$ for I_{g*}, g any element of G, we have proved most of the following statement.

(3.1) *The mapping of G into the group of all automorphisms of $\mathfrak{g}$ defined by $g \to \operatorname{Ad} g$ is a homomorphism. Let $Gl(\mathfrak{g})$ denote the group of all nonsingular linear transformations of $\mathfrak{g}$ as a vector space. Then $\operatorname{Ad}: G \to Gl(\mathfrak{g})$ is C^∞.*

It is only the last statement which requires proof and interpretation. In general, if V is a finite-dimensional vector space over $\boldsymbol{R}$, then the group $Gl(V)$ of all nonsingular linear transformations of V onto V is isomorphic to $Gl(n, \boldsymbol{R})$, $n = \dim V$. The isomorphism depends on the choice of a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V and is given by letting $A \in Gl(V)$ correspond to the matrix (α_{ij}) defined by $A(\mathbf{e}_j) = \sum_{i=1}^n \alpha_{ij} \mathbf{e}_i$, $j = 1, \ldots, n$. We take the topology and C^∞ structure on $Gl(V)$ obtained by identifying it with the Lie group $Gl(n, \boldsymbol{R})$. It may be shown (Exercise 3) that this C^∞ structure is independent of the choice of basis. Therefore, if we choose a basis $X_1, \ldots, X_n$ of $\mathfrak{g}$ and let $(\alpha_{ij}(g))$ denote the matrix corresponding in this way to $\operatorname{Ad} g$, the last statement asserts that $g \mapsto (\alpha_{ij}(g))$ is a C^∞ mapping. Note that $I_g(e) = e$, hence $I_{g*}: T_e(G) \to T_e(G)$. Because $\mathfrak{g}$ may be naturally identified with $T_e(G)$ by identifying each $X \in \mathfrak{g}$ with its value X_e at e, we may think of $\operatorname{Ad} g$ as a linear transformation on $\mathfrak{g}$—the left-invariant vector fields—or on $T_e(G)$ where, of course, it coincides with that induced by I_g according to the definition. In particular, if we use the latter point of view, the matrix $(\alpha_{ij}(g))$ is a submatrix of the Jacobian matrix evaluated at (g, e) of the C^∞ mapping of $G \times G \to G$ defined by $(g, x) \mapsto gxg^{-1} = I_g(x)$. Hence $g \mapsto (\alpha_{ij}(g))$ is C^∞. More generally, we make the following definitions.

(3.2) Definition A *representation* of a Lie group G on a vector space V is a Lie group homomorphism of G into the group $Gl(V)$ of nonsingular linear transformations of V onto V. Its *degree* (*dimension*) is the dimension of V. A *matrix representation* of G of degree n is a Lie group homomorphism of G into $Gl(n, \boldsymbol{R})$. The representation $g \mapsto \operatorname{Ad} g$ is called the *adjoint representation* of G.

We remark again that we interpret $\operatorname{Ad} g$ both as a linear mapping on $\mathfrak{g}$, the space of invariant vector fields, and on $T_e(G)$, the tangent space at the identity. This is by virtue of the identification of $\mathfrak{g}$ with $T_e(G)$. In either case $\operatorname{Ad} g$ is induced by the diffeomorphism $I_g(x) = gxg^{-1}$ of G onto G.

Many questions about Lie groups may be reduced to questions about the adjoint representation of the group. Some examples are given in the exercises; we give another below. First we shall need a definition and a lemma.

(3.3) Definition A covariant tensor field Φ of order r on G is *left-* (*right-*) *invariant* if $L_a^*\Phi_{ag} = \Phi_g$ (or $R_a^*\Phi_{ga} = \Phi_g$, respectively). It is *bi-invariant* if it is both left- and right-invariant.

We remark that any left- (or right-) invariant covariant tensor field $\Phi \in \mathscr{T}^r(G)$ is necessarily C^∞. If $X_1, \ldots, X_n$ is a basis of C^∞ left- (or right-) invariant vector fields, then $\Phi(X_{i_1}, \ldots, X_{i_r})$ is constant—hence C^∞—on G for any $1 \le i_1, \ldots, i_r \le n$. Therefore the components of Φ with respect to a C^∞-frame field are C^∞, and Φ is thus C^∞.

(3.4) Lemma *Let Φ_e be a covariant tensor of order r on the tangent space $T_e(G)$ at the identity. Then there is a unique left-invariant tensor field and a unique right-invariant tensor field coinciding at e with Φ_e. These two agree everywhere on G, that is, Φ_e determines a bi-invariant tensor field, if and only if* $\operatorname{Ad}(g)^*\Phi_e = \Phi_e$ *for all $g \in G$.*

Proof Let Φ_e be given on $T_e(G)$. For each $g \in G$ we have a unique left translation $L_g \colon G \to G$ which takes e to g. Define $\Phi \in \mathscr{T}^r(G)$ by $\Phi_g = L_{g^{-1}}^*\Phi_e$. Then $L_a^*\Phi_{ag} = L_a^*(L_{g^{-1}a^{-1}}^*\Phi_e) = L_a^* \circ L_{a^{-1}}^* \circ L_{g^{-1}}^*\Phi_e = L_{g^{-1}}^*\Phi_e$. However, this is just Φ_g, so we see that Φ is left-invariant. Similar arguments show that $R_{g^{-1}}^*\Phi_e$ is a right-invariant tensor field. Note that $R_{ab}^* = (R_b \circ R_a)^* = R_a^* \circ R_b^*$.

If Φ is bi-invariant, then $\operatorname{Ad}(g)^*\Phi_e = L_g^* \circ R_{g^{-1}}^*\Phi_e = \Phi_e$. Conversely, if this relation holds, then

$$L_{g^{-1}}^*\Phi_e = L_{g^{-1}}^* \circ L_g^* \circ R_{g^{-1}}^*\Phi_e = R_{g^{-1}}^*\Phi_e$$

so that the left- and right-invariant tensor fields determined by Φ_e agree at every $g \in G$. It is immediate that an invariant field must be determined by its value at any one element, say e, of G. ∎

(3.5) Corollary *Every Lie group has a left-invariant Riemannian metric and a left-invariant volume element. In particular every Lie group is orientable.*

Proof We take any inner product Φ_e (a positive definite, symmetric covariant tensor of order 2) on $T_e(G)$ and apply Lemma 3.4 to Φ_e and to the volume element Ω_e that Φ_e, with a choice of orientation of $T_e(G)$, determines in order to obtain a left-invariant Riemannian metric Φ and volume element Ω. ∎

In case G is compact we are able to say even more, as the next theorem and its corollary show. The corollary will make use of integration; to simplify the treatment we shall suppose G is connected (see Exercise 5).

(3.6) Theorem *An oriented, compact, connected Lie group G has a unique bi-invariant volume element Ω such that* vol $G = 1$.

Proof Let Ω be a left-invariant volume element on G. We claim that Ω is necessarily right-invariant also. In order to prove this it is enough to show that $\mathrm{Ad}(g)^*\Omega_e = \Omega_e$ for all $g \in G$. Let $X_1, \ldots, X_n$ be a basis of $\mathfrak{g}$ and X_{ie}, $i = 1, \ldots, n$, be the corresponding basis of $T_e(G)$. We have seen that $\mathrm{Ad}(g)X_j = \sum_{i=1}^n \alpha_{ij}(g)X_i$ and that $g \mapsto (\alpha_{ij}(g))$ defines a C^∞ homomorphism of $G \to Gl(n, \boldsymbol{R})$. The linear transformation $\mathrm{Ad}(g)^*$ on $\bigwedge^n(T_e(G))$ determined by $\mathrm{Ad}(g)$ acts as follows on Ω_e:

$$\mathrm{Ad}(g)^*\Omega_e = \det(\alpha_{ij}(g))\Omega_e .$$

However, since G is compact and connected, the same applies to its image under the C^∞-homomorphism $g \to \det(\alpha_{ij}(g))$ of G to R^*, the multiplicative group of nonzero real numbers. However, the only compact connected subgroup of R^* is $\{+1\}$, the trivial group consisting of the identity, hence $\det(\alpha_{ij}(g)) = 1$ and $\mathrm{Ad}(g)^*\Omega_e = \Omega_e$ for all $g \in G$. By the preceding lemma this proves that Ω is bi-invariant.

Any other bi-invariant Ω must be of the form $\lambda\Omega$, where λ is a positive constant; but then vol $G = \int_G \lambda\Omega = \lambda \int_G \Omega$. Hence it is possible to choose just one $\lambda \neq 0$ such that vol $G = +1$. For the opposite orientation on G, we would have $-\Omega$ as the corresponding unique bi-invariant volume element. ∎

From the existence of such a bi-invariant volume element one is able to deduce many important properties of Lie groups, of which the next two corollaries give examples. Further implications will appear later.

(3.7) Corollary *On a compact connected Lie group G it is possible to define a bi-invariant Riemannian metric $\tilde{\Phi}$.*

Proof Let Φ_e be a symmetric, positive definite, bilinear form on $T_e(G)$ and let Ω be the bi-invariant volume element. Given $X_e, Y_e \in T_e(G)$, we define a function on G by

$$f(g) = (\mathrm{Ad}(g)^*\Phi_e)(X_e, Y_e) = \Phi_e(\mathrm{Ad}(g)X_e, \mathrm{Ad}(g)Y_e),$$

the last equality being just the usual definition of $\mathrm{Ad}(g)^*$. Then define the bilinear form $\tilde{\Phi}_e$ on $T_e(G)$ by

$$\tilde{\Phi}_e(X_e, Y_e) = \int_G f(g)\Omega.$$

According to Lemma 3.4, $\tilde{\Phi}_e$ determines a bi-invariant form if for every $a \in G$

$$\mathrm{Ad}(a)^*\tilde{\Phi}_e(X_e, Y_e) = \tilde{\Phi}_e(X_e, Y_e).$$

The left-hand term may be written $\tilde{\Phi}_e(\mathrm{Ad}(a)X_e, \mathrm{Ad}(a)Y_e)$. Applying the definition of $\tilde{\Phi}_e$ to this expression, we find that

$$\begin{aligned}(\mathrm{Ad}(a))^*\tilde{\Phi}_e(X_e, Y_e) &= \int_G (\mathrm{Ad}(g)^*\Phi_e)\ (\mathrm{Ad}(a)X_e, \mathrm{Ad}(a)Y_e)\Omega \\ &= \int_G \mathrm{Ad}(a)^*\ \mathrm{Ad}(g)^*\Phi_e(X_e, Y_e)\Omega \\ &= \int_G \mathrm{Ad}(ga)^*\Phi_e(X_e, Y_e)\Omega,\end{aligned}$$

since $\mathrm{Ad}(a)^* \circ \mathrm{Ad}(g)^* = (\mathrm{Ad}(g) \circ \mathrm{Ad}(a))^* = \mathrm{Ad}(ga)^*$. This shows that

$$\mathrm{Ad}(a)^*\tilde{\Phi}(X_e, Y_e) = \int_G f(R_a(g))\Omega.$$

On the other hand, $R_a\colon G \to G$ is a diffeomorphism and Theorem 2.2 (iv) asserts that

$$\int_{R_a(G)} f(g)\Omega = \int_G f(R_a(g))R_a^*\Omega.$$

Since $R_a(G) = G$ and $R_a^*\Omega = \Omega$, we see that

$$\mathrm{Ad}(a)^*\tilde{\Phi}(X_e, Y_e) = \int_G f(g)\Omega = \tilde{\Phi}(X_e, Y_e).$$

It follows that $\tilde{\Phi}$ is a bi-invariant bilinear form on G. It is obviously symmetric and it is easy to check that it is positive definite. Since we do so in a more general case below, we will omit this verification here. ∎

(3.8) Remark When we use this Riemannian metric on G, we see that both right and left translations are isometries, that is, they preserve the

Riemannian metric (and also its associated distance function). We shall see later that as Riemannian manifolds compact Lie groups have rather interesting properties.

In closing this section we shall give another application, not too different from the above, which in fact actually includes it. Let (ρ, V), $\rho: G \to Gl(V)$, be a representation of G on a finite-dimensional real vector space V. As we have noted, if a basis is chosen in V, this determines a C^∞ homomorphism of G into $Gl(n, \boldsymbol{R})$, $n = \dim V$, a special case is $\rho = \mathrm{Ad}$ with $V = \mathfrak{g}$.

(3.9) Theorem *If G is compact and connected and ρ is a representation of G on V, then there is an inner product $(\mathbf{u}, \mathbf{v})$ on V such that every $\rho(g)$ leaves the inner product invariant:*

$$(\rho(g)\mathbf{u}, \rho(g)\mathbf{v}) = (\mathbf{u}, \mathbf{v}).$$

Proof Let $\Phi(\mathbf{u}, \mathbf{v})$ be an arbitrary inner product on V and, given a fixed $\mathbf{u}, \mathbf{v} \in V$, let $f(g) = \Phi(\rho(g)\mathbf{u}, \rho(g)\mathbf{v})$, thus defining a C^∞ function on G. Then we define

$$(\mathbf{u}, \mathbf{v}) = \int_G f(g)\Omega$$

with Ω denoting the bi-invariant volume element. The linearity of the integral implies at once that $(\mathbf{u}, \mathbf{v})$ is bilinear, and it is clearly symmetric in $\mathbf{u}, \mathbf{v}$ since the integrand is. Moreover $(\mathbf{u}, \mathbf{u}) \geq 0$, and equality implies $\mathbf{u} = 0$, by virtue of the fact that $f(g) \geq 0$ on G with equality holding if and only if the integral vanishes. Finally, for $a \in G$ we have

$$(\rho(a)\mathbf{u}, \rho(a)\mathbf{v}) = \int_G \Phi(\rho(g)\rho(a)\mathbf{u}, \rho(g)\rho(a)\mathbf{v})\Omega$$

$$= \int_G \Phi(\rho(ga)\mathbf{u}, \rho(ga)\mathbf{v})\Omega = \int_G f(ga)\Omega.$$

But by the same argument as in the previous proof, this is equal to $\int_G f(g)\Omega = (\mathbf{u}, \mathbf{v})$. This completes the proof. If we let $\rho = \mathrm{Ad}$ and $V = \mathfrak{g}$, we obtain Corollary 3.7 as a special case. ∎

We could state this result as follows: Each $\rho(g)$ is an *isometry* of the vector space V with the inner product $(\mathbf{u}, \mathbf{v})$. Since the matrix of an isometry of V relative to an orthonormal basis is an orthogonal matrix, we have the following corollary concerning the representations of a compact group.

(3.10) Corollary *Relative to a suitable basis of V, the matrices representing every $\rho(g)$ are orthogonal.*

Theorem 3.9 is very important in the representation theory of compact Lie groups. We shall say that $W \subset V$ is *invariant* if it is invariant for every linear transformation $\rho(g)$. The representation is *irreducible* if V contains no nontrivial invariant subspaces; if each invariant subspace W has a complementary invariant subspace W', such that $V = W \oplus W'$, then the representation is said to be *semisimple*. In this case it is easily verified that $V = W_1 \oplus \cdots \oplus W_r$, where the W_i are invariant irreducible subspaces. Applying Theorem 3.9 gives an important result.

(3.11) Corollary *If ρ is a representation of a compact connected Lie group G on a finite-dimensional vector space V, then it is semisimple. Moreover $V = W_1 \oplus \cdots \oplus W_r$, where for $i \neq j$ the subspaces are mutually orthogonal and each is a nontrivial irreducible subspace.*

Proof If V is irreducible, there is nothing to prove. If V contains a nontrivial invariant subspace W, then its orthogonal complement $W^\perp$ is also invariant: Let $\mathbf{w} \in W^\perp$ and let $\mathbf{v} \in W$. Then $(\rho(g)\mathbf{v}, \rho(g)\mathbf{w}) = (\mathbf{v}, \mathbf{w}) = 0$. Thus $\rho(g)\mathbf{w}$ is orthogonal to $\rho(g)\mathbf{v}$ for every $\mathbf{v} \in W$. Since $\rho(g)$ is nonsingular, this means that $\rho(g)\mathbf{w}$ is orthogonal to every element of W and must then be in $W^\perp$. Hence $V = W \oplus W^\perp$, a direct sum of complementary invariant subspaces. Repeated application of this argument gives the final statement of the corollary. ∎

(3.12) Example It is easy to see that there are representations of noncompact connected groups which do not have the property of complete reducibility, hence cannot leave an inner product invariant. For a simple example consider $\rho\colon \mathbf{R} \to Gl(2, \mathbf{R})$ acting on V^2 defined by

$$\rho(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Then $\rho(t)$ acts on V^2, the space of all $\binom{x}{y}$, $x, y \in \mathbf{R}$,

$$\rho(t)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + ty \\ y \end{pmatrix}.$$

The subspace $\binom{x}{0}$ is invariant but has no complementary invariant subspace.

Exercises

1. Show that a Lie group G has a bi-invariant volume element Ω if and only if it is possible to choose a basis $X_1, \ldots, X_n$ of $T_e(G)$ such that the matrix representation $g \to (\alpha_{ij}(g))$ corresponding to $\mathrm{Ad}(g)$ lies in $Sl(n, \mathbf{R})$, that is, $\det(\alpha_{ij}(g)) = +1$ for all $g \in G$. Give an example of such a G which is not compact.

2. Show that if the adjoint representation of G is irreducible, that is, there is no nontrivial subspace $V \subset T_e(G)$ which is invariant under $\mathrm{Ad}(g)$ for all g, then any normal subgroup $H \subset G$ has dimension 0 or n ($= \dim G$).
3. If V is a finite-dimensional vector space over $\boldsymbol{R}$ and $Gl(V)$ denotes the group of all nonsingular linear transformations of V onto V, then $Gl(V)$ is a group, isomorphic to $Gl(n, \boldsymbol{R})$, $n = \dim V$, with the isomorphism resulting from a choice of basis. Show that the C^∞ structure on $Gl(V)$ obtained from such an isomorphism is independent of this choice.
4. In Exercise 3 let $V = \mathfrak{g}$, the Lie algebra of a Lie group. Prove that the subgroup $\mathrm{Aut}\, \mathfrak{g}$, consisting of all elements of $Gl(\mathfrak{g})$ which are isomorphisms of $\mathfrak{g}$, is closed in $Gl(\mathfrak{g})$. Assume that each $\alpha \in \mathrm{Aut}\, \mathfrak{g}$ is given by F_*, where $F\colon G \to G$ is an automorphism and show that $\mathrm{Ad}(G)$ is a normal subgroup of $\mathrm{Aut}\, \mathfrak{g}$.
5. Show that the connected component of the identity G_0 in any Lie group G is an open and closed set and a normal (Lie) subgroup. Show also that if G_0 has a bi-invariant volume element or Riemannian metric, then the same is true of G if G is compact.

In the following three exercises let G be a connected Lie group and H a closed (Lie) subgroup of G and use the notation and ideas of Section IV.9. In particular, we denote by $\pi\colon G \to G/H$ the natural projection and by $\lambda\colon G \times G/H \to G/H$ the natural, transitive, left action of G on G/H.

6. Let o denote the coset H as a point of G/H and let λ' denote the action of H on G/H, obtained by restriction of λ to H. Show that for each $h \in H$, $\lambda'_h\colon G/H \to G/H$ leaves o fixed and that the correspondence $h \to \lambda'_{h*}$ is a representation of H on $T_0(G/H)$.
7. Using $\pi_*\colon T_e(G) \to T_0(G/H)$, show that $T_0(G/H)$ is naturally isomorphic to $\mathfrak{g}/\mathfrak{h}$ and that if these spaces are identified by this isomorphism, then the adjoint representation of G on $\mathfrak{g}$, when restricted to H, induces on $\mathfrak{g}/\mathfrak{h}$ the same representation as the one above.
8. Show that $M = G/H$ has a G-invariant Riemannian metric if H is compact.

4 Manifolds with Boundary

The problems we wish to consider when we deal with integration make it useful to introduce the notion of *manifold with boundary*, which we shall define presently. Examples are a line segment or ray, a circular disk or half-plane, a closed n-ball, a surface with an open disk removed, and so on. Manifolds with boundary are important for other reasons too, for example, to study differentiable deformations of differentiable maps from a manifold M to a manifold N, we will need to define C^∞ mappings from $M \times I$ into N. However, $M \times I$ is a manifold with boundary, for example, if $M = S^1$, then

$M \times I$ is a cylinder. What we must do, then, is extend our notions of differentiable functions and mappings, of tangent space and tensor field, and so on, to these slightly more general objects. In the definition of manifold with boundary the half-planes H^n play a role analogous to that of $\boldsymbol{R}^n$ for ordinary manifolds.

Let $H^n = \{x = (x^1, \ldots, x^n) \in \boldsymbol{R}^n \mid x^n \geq 0\}$ with the relative topology of $\boldsymbol{R}^n$, and denote by ∂H^n the subspace defined by $\partial H^n = \{x \in H^n \mid x^n = 0\}$. Then ∂H^n is the same space whether considered as a subspace of $\boldsymbol{R}^n$ or H^n; it is called the *boundary* of H^n. Of course all of these spaces carry the metric topology derived from the metric of $\boldsymbol{R}^n$, and ∂H^n is obviously homeomorphic to $\boldsymbol{R}^{n-1}$ by the map $(x^1, \ldots, x^{n-1}) \to (x^1, \ldots, x^{n-1}, 0)$.

Remembering now that differentiability has been defined for functions and mappings to $\boldsymbol{R}^m$ of *arbitrary* subsets of $\boldsymbol{R}^n$, we see that the notion of diffeomorphism applies at once to (relatively) open subsets U, V of H^n; namely, U, V are diffeomorphic if there exists a one-to-one map $F: U \to V$ (onto) such that F and F^{-1} are both C^∞ maps. Although this sounds precisely like the earlier definition, it is broader since U, V are not necessarily open subsets of $\boldsymbol{R}^n$, but are in fact the intersections of such sets with H^n. If $U, V \subset \boldsymbol{R}^n - \partial H^n$, then U and V are actually open in $\boldsymbol{R}^n$ so that this definition of diffeomorphism coincides with our previous one. On the other hand, if $U \cap \partial H^n \neq \varnothing$, then we claim that $V \cap \partial H^n \neq \varnothing$ and that $F(U \cap \partial H^n) \subset V \cap \partial H^n$. Similarly, $F^{-1}(V \cap \partial H^n) \subset U \cap \partial H^n$; in other words, diffeomorphisms on open sets of H^n take boundary points to boundary points and interior points to interior points. This follows at once from the inverse function theorem: $U - \partial H^n$ is open in $\boldsymbol{R}^n$ and hence F must map it diffeomorphically onto an open subset of $\boldsymbol{R}^n$, but no open subset of H^n which contains a boundary point, that is, a point of ∂H^n, can be open in $\boldsymbol{R}^n$. Thus $F(U - \partial H^n) \subset V - \partial H^n$ and $F^{-1}(V - \partial H^n) \subset U - \partial H^n$. Since F and F^{-1} are one-to-one on U and V, the result follows.

We also notice the following two facts: First $U \cap \partial H^n$ and $V \cap \partial H^n$ are open subsets of ∂H^n, a submanifold of $\boldsymbol{R}^n$ diffeomorphic to $\boldsymbol{R}^{n-1}$; and F, F^{-1} restricted to these open sets in ∂H^n are diffeomorphisms. Second both F and F^{-1} can be extended to open sets U', V' of $\boldsymbol{R}^n$ having the property that $U = U' \cap H^n$ and $V = V' \cap H^n$. These extensions will not be unique nor are the extensions in general inverses throughout these larger domains. However, the derivatives of F and F^{-1} on U and V are independent of the extensions chosen and we may suppose that even on the extended domains the Jacobians are of rank n. These statements are immediate consequences of the definition of differentiability for arbitrary subsets of $\boldsymbol{R}^n$ and the fact that the Jacobian of a C^∞ mapping has its maximum rank on an open subset of its domain. Some further amplification of this situation is given in the problems. The reader will find a complete discussion in Guillemin and Pollack [1].

(4.1) Definition *A C^∞ manifold with boundary* is a Hausdorff space M with a countable basis of open sets and a differentiable structure $\mathscr{U}$ in the following (generalized) sense (compare Definition III.1.2): $\mathscr{U} = \{U_\alpha, \varphi_\alpha\}$ consists of a family of open subsets U_α of M each with a homeomorphism φ_α onto an open subset of H^n (topologized as a subspace of $\boldsymbol{R}^n$) such that:

(1) the U_α cover M;
(2) if U_α, φ_α and U_β, φ_β are elements of $\mathscr{U}$, then $\varphi_\beta \circ \varphi_\alpha^{-1}$ and $\varphi_\alpha \circ \varphi_\beta^{-1}$ are diffeomorphisms of $\varphi_\alpha(U \cap V)$ and $\varphi_\beta(U \cap V)$, open subsets of H^n;
(3) $\mathscr{U}$ is maximal with respect to properties (1) and (2).

Examples are shown in Fig. VI.5.

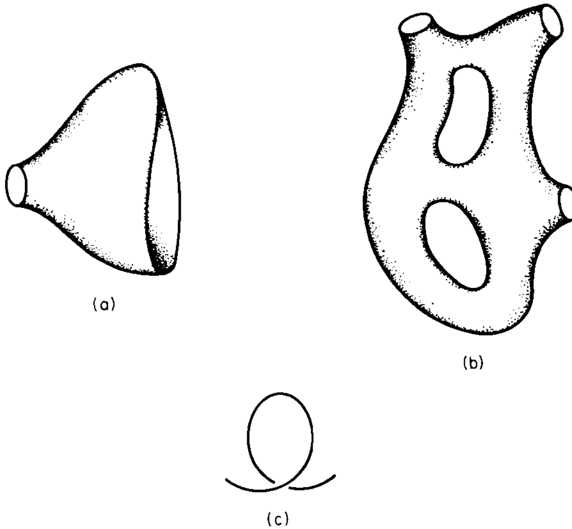


Figure VI.5

The U, φ are coordinate neighborhoods on M. From the remarks above we see that if $\varphi(p) \in \partial H^n$ in one coordinate system, then this holds for all coordinate systems. The collection of such points is called the *boundary* of M, denoted ∂M, and $M - \partial M$ is a manifold (in the ordinary sense), which we denote by Int M. If $\partial M = \varnothing$, then M is a manifold of the familiar type; we call it a manifold *without boundary* when it is necessary to make the distinction. The following theorem follows from the first of the two facts remarked upon above.

(4.2) Theorem *If M is a C^∞ manifold (of dimension n) with boundary, then the differentiable structure of M determines a C^∞-differentiable structure of dimension $n - 1$ on the subspace ∂M of M. The inclusion $i: \partial M \to M$ is an imbedding.*

The differentiable structure $\tilde{\mathscr{U}}$ on ∂M is determined by the coordinate neighborhoods $\tilde{U}, \tilde{\varphi}$, where $\tilde{U} = U \cap \partial M$, $\tilde{\varphi} = \varphi|_{U \cap \partial M}$ for any coordinate neighborhood U, φ of the atlas $\mathscr{U}$ of M which contains points of ∂M.

Differentiable functions, differentiable mappings, rank, and so on, may now be defined on M exactly as before by using local coordinates. By virtue of the C^∞ compatibility of such coordinate systems these concepts are independent of the choice of coordinates. We leave the verification to the reader. We also define $T_p(M)$ at boundary points of M. This could be done using derivations on $C^\infty(p)$ as before, but to avoid some slight complications we use an alternative definition. First note that in the case of $H^n \subset \boldsymbol{R}^n$, upon which manifolds with boundary are modeled, we identify $T_x(H^n)$ with $T_x(\boldsymbol{R}^n)$; we may think of this identification as being given by the inclusion mapping. For $x \in \partial H^n$, this defines what we mean by $T_x(H^n)$. In the case of a general manifold M, for $p \in \partial M$ we define a *vector* $X_p \in T_p(M)$ to be an assignment to each coordinate neighborhood U, φ of an n-tuple of numbers $(\alpha^1, \ldots, \alpha^n)$, the U, φ components of X_p satisfying the following condition: If $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ are coordinates around p in neighborhoods U, φ and V, ψ, then the components $(\alpha^1, \ldots, \alpha^n)$ and $(\beta^1, \ldots, \beta^n)$ relative to U and V are related by

$$\beta^i = \sum_{j=1}^{n} \left(\frac{\partial y^i}{\partial x^j}\right)_{\varphi(p)} \alpha^j, \qquad i = 1, \ldots, n$$

(as in Corollary IV.1.8). What this does is attach to each $p \in M$ a $T_p(M)$ such that each coordinate system U, φ determines an isomorphism φ_* taking X_p with components $(\alpha^1, \ldots, \alpha^n)$ to the vector $\sum \alpha^i(\partial/\partial x^i) \in T_{\varphi(p)}(H^n)$. As previously $E_1, \ldots, E_n$ will denote the basis determined by $\varphi_*(E_i) = \partial/\partial x^i$, $i = 1, \ldots, n$. Having defined $T_p(M)$ on ∂M [it is already known on Int M, which is an ordinary manifold], we may extend all of our definitions and theorems to manifolds with boundary. In particular, exterior differential forms and the exterior calculus is still valid on manifolds with boundary. There is no essential change in the definitions or proofs.

For many purposes, in particular for our discussion of Stokes's theorem in the next section, we could use an (apparently) weaker, but closely related notion.

(4.3) Definition A *regular domain* D on a manifold M is a closed subset of M with nonempty interior $\mathring{D}$ such that if $p \in \partial D = D - \mathring{D}$, then p has a cubical coordinate neighborhood U, φ with $\varphi(p) = (0, \ldots, 0)$, $\varphi(U) = C_\varepsilon^n(0)$, and $\varphi(U \cap D) = \{x \in C_\varepsilon^n(0) \mid x^n \geq 0\}$; thus $x^n = 0$ on ∂D.

It follows that ∂D is a regular submanifold. We remark that if D is compact, then it is a domain of integration on M. It is a straightforward matter to check that D, with the topology and differentiable structure induced

by M is a manifold with boundary. All of our examples can be seen to be of this type: H^n and the closed unit ball $\bar{B}^n$ are regular domains of $M = \boldsymbol{R}^n$, $N \times I$ is a regular domain of $N \times \boldsymbol{R}$, and the set D obtained by removing from a manifold M a diffeomorphic image of an open ball is a regular domain. Further examples are given in the exercises.

It is a fact—somewhat too difficult to prove here—that any manifold M with boundary can be realized as a regular domain of a larger manifold M'. The basic idea is simple enough: one simply takes two copies of M, say M_1 and M_2, and "glues" them together along their boundaries, identifying corresponding boundary points. The resulting manifold, called the *double of* M, contains M as a regular domain. Figure VI.6 shows the doubles of the examples of Fig. VI.5 (which are shaded in Fig. VI.6). For details the reader

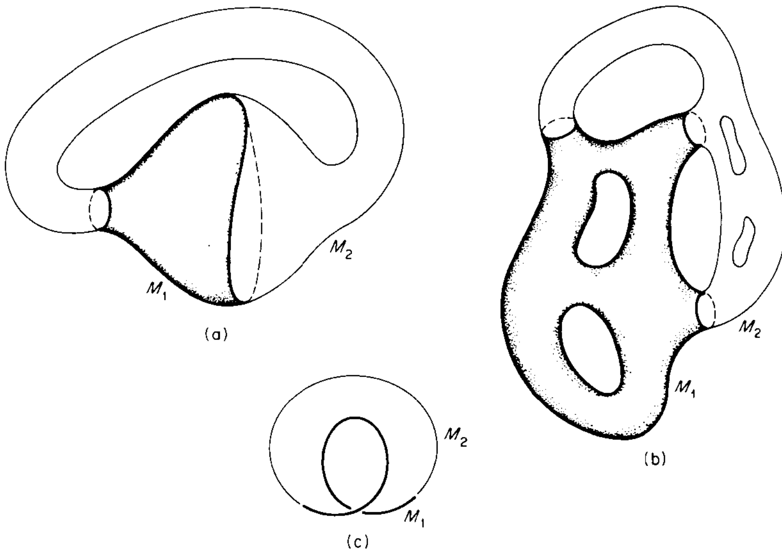


Figure VI.6

should consult Munkres [1, Section 6]. For regular domains it is simpler or at least more intuitive to define the tangent space at boundary points and to define the calculus of exterior differential forms, since we may do so by restriction of the corresponding objects on M. This could be taken as further evidence that the reader who wishes to carefully check the details of the extension to manifolds with boundary of the concepts and operations we have used for manifolds without boundary, will encounter no serious obstacle. In fact, by definition M is *locally* diffeomorphic to H^n, which is a regular domain of $\boldsymbol{R}^n$.

We consider next the question of orientability. A manifold M with non-empty boundary is orientable provided that it has a covering of coordinate

neighborhoods $\{U_\alpha, \varphi_\alpha\}$ which are coherently oriented, that is, $U_\alpha \cap U_\beta \neq \varnothing$ implies $\varphi_\beta \circ \varphi_\alpha^{-1}$ has positive Jacobian determinant (or equivalently, preserves the natural orientation of H^n). This is equivalent to the existence of a nowhere vanishing n-form Ω on M. The proof is the same except that when we speak of a partition of unity on M associated to a regular covering $\{U_i, V_i, \varphi_i\}$ we limit ourselves to a regular covering by cubical coordinate neighborhoods concerning which we impose the following slight restriction: if $U_i \cap \partial M \neq \varnothing$, then $\varphi_i(U_i) = C_3^n(0) \cap H^n$ and $\varphi_i(V_i) = C_1^n(0) \cap H^n$. With this modified definition of regular covering we still have a regular covering (by definition locally finite) refining any open covering $\{A_\alpha\}$ of M and an associated C^∞ partition of unity $\{f_i\}$ on M. We remark that those U_i, V_i, φ_i of the regular covering that intersect ∂M determine a regular covering $\tilde{U}_i = U_i \cap \partial M$, $\tilde{V}_i = V_i \cap \partial M$, and $\tilde{\varphi}_i = \varphi_i|_{\tilde{U}_i}$ of ∂M and the associated partition of unity restricts to an associated partition of unity $\{\tilde{f}_i = f_i|_{\partial M}\}$ on ∂M.

(4.4) Theorem *Let M be an oriented manifold and suppose ∂M is not empty. Then ∂M is orientable and the orientation of M determines an orientation of ∂M. In particular, this holds for ∂D if D is a regular domain on M.*

Proof Since ∂M is an $(n-1)$-dimensional submanifold of M, its tangent space at each point may be identified with an $(n-1)$-dimensional subspace of $T_p(M)$; we denote this subspace by $T_p(\partial M)$. We shall show that there is a distinction between the two half-spaces into which $T_p(\partial M)$ divides $T_p(M)$ which is independent of coordinates. Suppose that U, φ and V, ψ are coordinate neighborhoods of $p \in \partial M$ with local coordinates $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$, respectively. By our definitions of coordinates of boundary points, the last coordinate x^n, or y^n is equal to zero if the point in U or V, respectively, is on ∂M, and positive otherwise. Writing $y^i = y^i(x^1, \ldots, x^n)$, $i = 1, \ldots, n$, for the change of coordinate functions, we have $0 = y^n(x^1, \ldots, x^{n-1}, 0)$ so that $(\partial y^n/\partial x^1)_{\varphi(q)} = \cdots = (\partial y^n/\partial x^{n-1})_{\varphi(q)} = 0$ for every $q \in U \cap \partial M$. It follows that the Jacobian matrix then has the form

$$D(\psi \circ \varphi^{-1}) = \begin{pmatrix} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial y^1}{\partial x^{n-1}} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^{n-1}} & 0 \\ \dfrac{\partial y^1}{\partial x^n} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^n} & \dfrac{\partial y^n}{\partial x^n} \end{pmatrix}_{\varphi(q)}$$

Since the Jacobian is nonsingular, $\partial y^n/\partial x^n \neq 0$ at $\varphi(q)$; in fact, it must be positive. For let $\varphi(q) = (a^1, a^2, \ldots, a^{n-1}, 0)$ and consider $f(t)$, defined by

$f(t) = y^n(a^1, \ldots, a^{n-1}, t)$. Then $f'(0) = (\partial y^n/\partial x^n)_{\varphi(q)}$ can certainly not be negative since $f(0) = 0$ and $f(t) > 0$ in some interval $0 < t < \delta$; therefore $\partial y^n/\partial x^n > 0$ at $\varphi(q)$ as claimed.

If U, φ and V, ψ are *oriented* neighborhoods of M, then $D(\psi \circ \varphi^{-1})$ has positive determinant so both $\partial y^n/\partial x^n$ and the $(n-1) \times (n-1)$ minor determinant obtained by striking out the last row and column have the same sign. This minor determinant is exactly the determinant of $D(\tilde{\psi} \circ \tilde{\varphi}^{-1})$, the change of coordinates from $\tilde{U} = U \cap \partial M$, $\tilde{\varphi} = \varphi|_{\tilde{U}}$ to $\tilde{V} = V \cap \partial M$, $\tilde{\psi} = \psi|_{\tilde{V}}$ on the submanifold ∂M. Thus the neighborhoods on ∂M determined by oriented neighborhoods on M are coherent and determine an orientation on ∂M.

(4.5) Remark Using the notation of the proof, let $q \in U \cap V$ be a boundary point of M and let $X_q \in T_q(M)$. Because $(\partial y^n/\partial x^m)_{\varphi(q)} > 0$, it follows that when we express X_q in the coordinate frames of either U, φ or V, ψ,

$$X_q = \alpha^1 E_1 + \cdots + \alpha^{n-1} E_{n-1} + \alpha^n E_n = \beta^1 F_1 + \cdots + \beta^{n-1} F_{n-1} + \beta^n F_n,$$

then α^n and β^n have the same sign. (This fact does *not* depend on the coordinates being oriented.) It follows that the vectors of $T_p(M) - T_p(\partial M)$ fall into two classes, those whose last component is positive—which we call *inward pointing vectors* at $p \in \partial M$—and those for which the last component is negative—which we call *outward pointing vectors* (see Fig. VI.7). Those for which the last component vanishes are tangent to ∂M. This classification is independent of the orientation of M. The orientation above on ∂M could be described as follows: If $p \in \partial M$ and X_p is an inward pointing vector, then a basis $E_{1p}, \ldots, E_{(n-1)p}$ of $T_p(\partial M)$ is oriented if and only if $E_{1p}, \ldots, E_{(n-1)p}, X_p$ is an oriented basis of $T_p(M)$. Whenever M is oriented, ∂M will denote this orientation on the boundary. For Stokes's Theorem (below) it is convenient

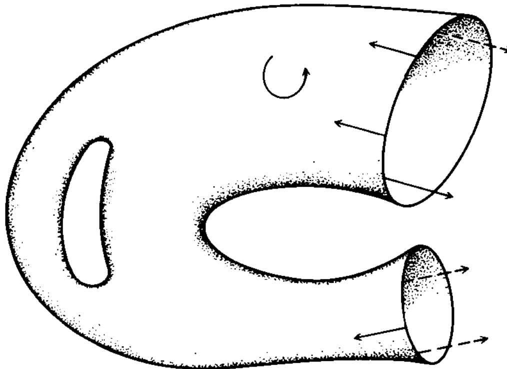


Figure VI.7

to use the *opposite* orientation $-\partial M$ when n is odd. We will denote this orientation on the boundary by $\tilde{\partial} M$. Thus $\tilde{\partial} M = (-1)^n \partial M$

(4.6) Remark We have noted that there are difficulties in gluing two manifolds with identical boundaries together along their boundaries. We can, however, describe a special case which will give some idea of the importance of such operations. Let M_1, M_2 be two manifolds (without boundary) of dimension n and let U_i, φ_i be coordinate neighborhoods of points $p_i \in M_i$, $i = 1, 2$. We suppose that $\varphi_i(p_i) = (0, \ldots, 0)$ and that $\varphi_i(U_i) = B_2^n(0)$ in each case and we set $V_i = \varphi_i^{-1}(B_1(0))$. Then $M_i' = M_i - V_i$, $i = 1, 2$, is a manifold with boundary, indeed $\varphi_i(\partial M_i') = S^{n-1}$. The manifold obtained by gluing M_1' to M_2' along the boundaries is called the *connected sum* of M_1 and M_2, denoted $M_1 \# M_2$ (see Fig. VI.8). In order to define it without loss of differentiability, we will actually remove only $\varphi^{-1}(\bar{B}_{1/2}(0))$ from each M_i to obtain M_i'', leaving in each U_i a "collar," $C_i = \varphi_i^{-1}\ (B_{3/2}(0) - \bar{B}_{1/2}(0))$ parametrized by $\varphi_i^{-1}\{(x, t) = (x^1, \ldots, x^n, t) \mid \|x\| = 1,\ \frac{1}{2} < t < \frac{3}{2}\}$, a diffeomorph of $S^{n-1} \times (\frac{1}{2}, \frac{3}{2})$. We identify points of C_1 corresponding to (x, t) with those of C_2 corresponding to $(x, 2 - t)$.

Any closed surface (compact 2-manifold) can be obtained as the connected sum of copies of S^2 and T^2 if orientable, $P^2(R)$ and T^2 if nonorientable (see Wallace [1], Massey [1], and Hirsch [1]).

This whole procedure and ones similar to it have become very important in the recent years and are intimately related to the attempt to classify or list all simply connected n-dimensional compact C^∞ manifolds—a problem which was solved long ago for closed surfaces but is still unsettled in dimension three. Oddly enough there has been more success in higher dimensions! It is not yet known whether there exist simply connected, compact, orientable manifolds of dimension three other than S^3; that there are none is the famous Poincaré conjecture. Similar questions in dimension ≥ 5 were

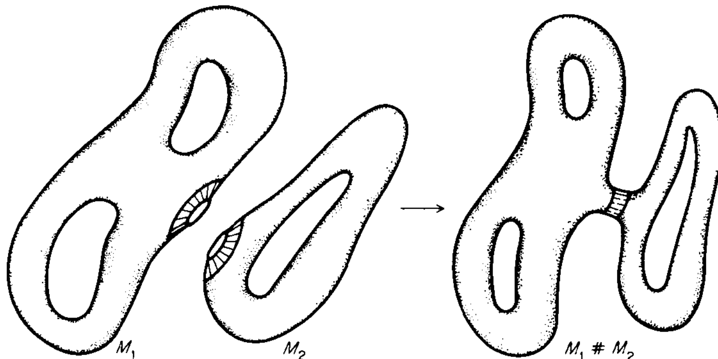


Figure VI.8

answered by Smale [1]. Milnor [3] has shown that every compact 3-manifold can be represented uniquely as a connected sum of 3-manifolds which cannot be further decomposed into connected sums.

Exercises

1. Let M be a manifold without boundary and $f\colon M \to \boldsymbol{R}$ a C^∞ function. Suppose that $df \neq 0$ on $f^{-1}(0)$ and show that $M^+ = \{p \in M \mid f(p) \geq 0\}$ is a regular domain of M. What is its boundary?
2. Let C be an imbedded image of S^1 in $\boldsymbol{R}^3$. Show that there is an $\varepsilon > 0$ such that $N = \bigcup_{x \in C} \bar{B}_\varepsilon(x)$ is a manifold with boundary. Show that it is diffeomorphic to the solid torus. What can be said about the complement in $\boldsymbol{R}^3$ of $\mathring{N}$ (the interior of N)?
3. Show that if M has a Riemannian metric, then there is a uniquely determined vector field X defined at each point of ∂M such that X_p is inward pointing, is orthogonal to $T_p(\partial M)$ for each $p \in \partial M$, and has unit length.
4. Show that if M is a manifold with boundary, then it is always possible to choose a vector field X defined at each point of ∂M such that X is inward pointing. Given such X and an Ω on M which is an n-form determining the orientation, then show that the form $\omega = (-1)^{n-1} i(X)\Omega$ determines the orientation of ∂M and $i(X)\Omega$ that of $\tilde{\partial} M$. [$i(X)\Omega$ is defined as in the Exercise of Section V.8.4].
5. Let M be a connected, closed, regular 2-manifold in $\boldsymbol{R}^3$ such that $\boldsymbol{R}^3 - M$ is not connected. Show that it is possible to define a continuous field of unit normal vectors to M and that $\boldsymbol{R}^3 - M$ has two components, each with M as boundary.
6. Let U, V be open subsets of H^n and $F\colon U \to V$, $G\colon V \to U$ diffeomorphisms which are inverse to each other. Discuss the possibility of finding extensions F', G' to open subsets U', V' of $\boldsymbol{R}^n$ containing U, V, respectively, such that $G' \circ F' = id_{U'}$ and $F' \circ G' = id_{V'}$.

5 Stokes's Theorem for Manifolds with Boundary

We consider an oriented manifold M with possibly nonempty boundary ∂M, oriented by the orientation of M. We shall consider only *oriented* coordinate neighborhoods U, φ in what follows. If $U \cap \partial M \neq \varnothing$, then we denote by $\tilde{U}$, $\tilde{\varphi}$ the corresponding neighborhood $\tilde{U} = U \cap \partial M$, $\tilde{\varphi} = \varphi|_{\tilde{U}}$ on ∂M. All of the concepts used in defining the integral extend to M; namely the definitions of content zero, domain of integration, and so on. In particular ∂M has measure zero and, if compact, has content zero. This follows from corresponding properties of ∂H^n (and Corollary 1.14). A cube Q is associated

with U, φ as in Section 2 unless $U \cap \partial M \neq \varnothing$, in which case we assume that Q has a "face" on ∂M, that is,

$$\varphi(Q \cap \partial M) = \{x \in \mathbf{R}^n \mid 0 \le x^i \le 1 \quad \text{and} \quad x^n \equiv 0\}.$$

In this case we note two facts: (a) $\tilde{Q} = Q \cap \partial M$ is a cube of ∂M associated with $\tilde{U}$, $\tilde{\varphi}$ and (b) $\mathring{Q} = \varphi^{-1}(\{x \in \mathbf{R}^n | 0 < x^i < 1, 1 \le i \le n-1; 0 \le x^n < 1\})$, that is, the interior of Q has a different image in $\mathbf{R}^n$ than it has when $U \subset \operatorname{Int} M$.

Taking these minor modifications into account, the definition of $\int_M \Omega$ is exactly as in Section 2 and the integral of an integrable n-form has the same properties as before. Indeed, if M is a compact regular domain in a manifold N, then it is necessarily a domain of integration in N and $\int_M \Omega = \int_N k_M \Omega$ so there is nothing new to define in this case! The same comments apply to the integral over a Riemannian manifold with boundary and to the definition of vol M when M is compact.

Now suppose M is both oriented and compact and that ω is an $(n-1)$ form of class C^1 at least on M. We have an important relation between the integral of $d\omega$ over M and $i^*\omega$, the restriction of ω to ∂M ($i\colon \partial M \to M$ is the inclusion mapping). To simplify the statement of the theorem we let $\tilde{\partial} M$ denote ∂M, the boundary with the orientation induced by M, when n is even and $-\partial M$, the boundary with the opposite orientation when n is odd; thus $\tilde{\partial} M = (-1)^n\, \partial M$ (cf. Remark 4.5).

(5.1) Theorem (Stokes's theorem) *Let M be an oriented compact manifold of dimension n or a compact regular domain of an oriented manifold and let ∂M have the orientation $\tilde{\partial} M$ described above. Then we have*

$$\int_M d\omega = \int_{\tilde{\partial} M} i^*\omega.$$

When $\partial M = \varnothing$, the integral over M vanishes.

Proof According to our definitions it is enough to establish the theorem for an ω whose support is contained in the interior $\mathring{Q}$ of a cube Q associated to a coordinate neighborhood U, φ. Suppose ω has its support in Q and $x^1, \ldots, x^n$ are the local coordinates. We may suppose that in these coordinates ω is expressed as

$$\varphi^{-1*}(\omega) = \sum_{j=1}^{n} (-1)^{j-1} \lambda^j \, dx^1 \wedge \cdots \wedge dx^{j-1} \wedge dx^{j-1} \wedge \cdots \wedge dx^n.$$

Then we have

$$\varphi^{-1*}(d\omega) = d\varphi^{-1*}(\omega) = \left(\sum_{j=1}^{n} \frac{\partial \lambda^j}{\partial x^j} \right) dx^1 \wedge \cdots \wedge dx^n$$

so that

$$\int_M d\omega = \int_C \left(\sum_{j=1}^n \frac{\partial \lambda^j}{\partial x^j} \right) dv = \sum_j \int_0^1 \cdots \int_0^1 \frac{\partial \lambda^j}{\partial x^j} dx^1 \cdots dx^n.$$

This follows from the definition of integration on M and the iterated integral theorem. The expression on the right may be rewritten; consider the jth summand only and integrate first with respect to the variable x^j. This gives an $(n-1)$-fold iterated integral

$$(*) \qquad \int_0^1 \cdots \int_0^1 [\lambda^j(x^1, \ldots, x^{j-1}, 1, x^{j+1}, \ldots, x^n) - \lambda^j(x^1, \ldots, x^{j-1}, 0, x^{j-1}, \ldots, x^n)] \, dx^1 \cdots \widehat{dx^j} \cdots dx^n,$$

where we indicate by $\widehat{dx^j}$ that this differential is to be omitted. The sum of these $(n-1)$-fold iterated integrals for $j = 1, \ldots, n$ gives $\int_M d\omega$ if $\operatorname{supp}(\omega) \subset \mathring{Q}$. Two cases can occur: either $Q \cap \partial M = \varnothing$, in which case $\varphi(\mathring{Q}) = \{x \mid 0 < x^i < 1, \; i = 1, \ldots, n\}$; or $Q \cap \partial M \neq \varnothing$, in which case $\varphi(\mathring{Q}) = \{x \mid 0 < x^i < 1, i = 1, \ldots, n-1; 0 \leq x^n < 1\}$. In the first case, using $\operatorname{supp} \omega \subset \mathring{Q}$, we see that $\lambda^j = 0$ if any $x^j = 0, 1$. Hence each of the integrands in $(*)$ vanish and $\int_M d\omega = 0$. On the other hand ω restricted to ∂M is the zero $(n-1)$-form since $\operatorname{supp} \omega \subset \mathring{Q}$ which has no points on ∂M. Thus $\int_M d\omega = 0 = \int_{\partial M} i^*\omega$ and Stokes's theorem holds for this case.

In the second case we again have all of the integrands in $(*)$ equal to zero except the one corresponding to $j = n$; therefore

$$\int_M d\omega = -\int_0^1 \cdots \int_0^1 \lambda^n(x^1, \ldots, x^{n-1}, 0) \, dx^1 \cdots dx^{n-1}.$$

On the other hand we may evaluate $\int_{\partial M} i^*\omega$ using the fact that $i^*\omega$ has its support in $\tilde{Q} = Q \cap \partial M$ so that its expression in the local coordinates $\tilde{U}, \tilde{\varphi}$ (obtained by restriction of U, φ) collapses to

$$\tilde{\varphi}^{-1*}(i^*\omega) = (-1)^{n-1} \lambda^n(x^1, \ldots, x^{n-1}, 0) \, dx^1 \wedge \cdots \wedge dx^{n-1}.$$

[We may obtain this from the expression for ω by applying the corresponding inclusion $i: (x^1, \ldots, x^{n-1}) \to (x^1, \ldots, x^{n-1}, 0)$ in the local coordinates and noting that $i^* \, dx^n = 0$.] This will give

$$\int_{\partial M} i^*\omega = (-1)^{n-1} \int_0^1 \cdots \int_0^1 \lambda^n(x^1, \ldots, x^{n-1}, 0) \, dx^1 \cdots dx^{n-1}.$$

Thus, in the case where $\operatorname{supp} \omega \subset \mathring{Q}$ and $\mathring{Q} \cap \partial M \neq \varnothing$, we find that

$$\int_M d\omega = (-1)^n \int_{\partial M} i^*\omega = \int_{\pm \partial M} i^*\omega,$$

with the right-hand integral over ∂M when n is even and $-\partial M$ when n is odd, that is, over ∂M. ∎

When $n = 1$, $M = [a, b] \subset \boldsymbol{R}$ and Stokes's theorem is the fundamental theorem of calculus. Green's theorem and Gauss's theorem are also special cases, giving some idea of its great importance.

(5.2) Example (*Green's theorem*) Let M be a bounded regular domain of $\boldsymbol{R}^2$, that is, the closure of a bounded open subset of the plane bounded by simple closed curves of class C^∞; for example, let M be a circular disk or annulus. Then ∂M is the union of these curves (in the cases mentioned a circle or a pair of concentric circles). If ω is a one-form of class C^1 on M, then, using the natural Cartesian coordinates, we have $\omega = a\,dx + b\,dy$. We may suppose, by definition of differentiability on arbitrary sets, that a, b are restrictions of C^1 functions on some open set containing M. We have $d\omega = ((\partial b/\partial x) - (\partial a/\partial y))\,dx \wedge dy$ and by Stokes's theorem

$$\int_M \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right) dx \wedge dy = \int_{\partial M} a\,dx + b\,dy.$$

According to Remark 2.3, the left-hand side is the ordinary Riemann integral over the domain of integration $M \subset \boldsymbol{R}^2$. On the other hand, if we think of ∂M as a one-dimensional manifold and cover it with (oriented) neighborhoods, it is clear that its value is that of the usual line integral along a curve C (or curves C_i) oriented so that as we traverse the curve the region is on the left. (This is further discussed below.) Thus the equality above may be written

$$\iint_M \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right) dx\,dy = \sum_i \int_{C_i} a\,dx + b\,dy,$$

which is the usual statement of Green's theorem.

(5.3) Example Let M be a regular domain of $\boldsymbol{R}^3$, that is, the closure of a bounded open set bounded by closed C^∞ surfaces. Examples are the ball of radius 1, which is bounded by the sphere S^2, or the region interior to a torus T^2, obtained by rotating a circle around a line exterior to it. Consider the two-form $\omega = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$, where P, Q, R are C^1 functions on some open set of $\boldsymbol{R}^3$ containing M. We have

$$d\omega = \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz$$

and Stokes's theorem asserts that

$$\int_M \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz = \int_{-\partial M} P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy.$$

If we use Remark 2.3 and our definitions to translate this into a Riemann integral over a domain and a surface integral over the boundary, respectively, then we obtain the usual *divergence theorem* of advanced calculus.

(5.4) Example A third example is obtained if we consider M to be a piece of surface imbedded in $\boldsymbol{R}^3$ and bounded by smooth simple closed curves, for example, a sphere with one or more open circular disks removed, thus leaving boundary circles, which are ∂M. Since dx, dy, and dz may be considered, by restriction, as one-forms on M or on ∂M, any one-form ω on M may be written: $\omega = A\,dx + B\,dy + C\,dz$, where A, B, and C are C^1 functions on M. In this case Stokes's theorem asserts that

$$\int_M \left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}\right) dy \wedge dz + \left(\frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}\right) dz \wedge dx + \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx \wedge dy$$
$$= \int_{\partial M} A\,dx + B\,dy + C\,dz.$$

The integral on the left-hand side may be converted by the procedure used in defining integrals of forms on manifolds to an ordinary surface integral over the surface M in $\boldsymbol{R}^3$ and that on the right to a line integral. When this is done one obtains the usual Stokes theorem of advanced calculus.

Often these examples of special cases of the general Stokes theorem are stated in terms of vector calculus and vector operations such as *gradient*, *divergence*, and *curl*. To establish the equivalence would require use of the duality between vectors and covectors and other such relations on $\boldsymbol{R}^n$ which use the fact that $\boldsymbol{R}^n$ is a Riemannian manifold. Basically these stem from the natural isomorphism of a Euclidean vector space and its dual (and its extension to a duality between covariant and contravariant tensors). We do not need these operations in what follows so they will not be taken up here; some indications are given in the exercises.

It is important to note that the version of Stokes's theorem proved above is deficient in the following sense: it holds only for smooth manifolds with smooth boundary. Thus, for example, our proof does not even include the case of a square in $\boldsymbol{R}^2$ or an open set of $\boldsymbol{R}^3$ bounded by a polyhedron. The difficulty in these cases is not so much with the analysis and integration theory, as with describing the regions of integration to be admitted and with giving precise definitions of orientability and induced orientation of the boundary. The search for reasonable domains of integration to validate Stokes's theorem usually leads to the concept of a simplicial or polyhedral complex, that is, a space made up by fastening together along their faces a number of simplices (line segments, triangles, tetrahedra, and their generalizations) (Fig. VI.4) or more general polyhedra (cubes, for example). Since it can be shown (see Munkres [1]) that any C^∞ manifold M may be "triangulated," which means that it is homeomorphic (even with considerable smoothness) to such a complex, the integral over M becomes the sum of the integrals over the pieces, which are images of simplices, cubes, or other polyhedra as the case may be (compare Remark 2.7). The strategy is then to reduce the theory (including Stokes's theorem) to the case of polyhedral

domains of $\boldsymbol{R}^n$. This approach is particularly important for those interested in algebraic topology and de Rham's theorem. It is very clearly set forth, for example, by Singer and Thorpe [1] or Warner [1].

For many purposes, integration of forms over C^1 (but not one-to-one or diffeomorphic) images of simplices and polyhedra in a manifold is very useful. Some idea of how this works is given in the following example.

(5.5) Example (*Line integrals in a manifold*) Let $[a, b] = t \in \boldsymbol{R} | a \leq t \leq b\}$ and let $F: [a, b] \to M$ be a C^1 mapping whose image is, then, a C^1 curve S on M. If ω is a one-form on M, we define $\int_S \omega$ by

$$\int_S \omega = \int_{[a,b]} F^*\omega.$$

This is called the *line integral* of ω along S. In general S is not a submanifold of M; it can be very complicated. However, the right-hand side is the integral of a one-form, $F^*\omega = f(t)\,dt$, on a one-dimensional manifold with boundary; thus

$$\int_S \omega = \int_a^b f(t)\,dt.$$

Exactly as for line integrals in $\boldsymbol{R}^n$, we may prove that the value of the integral does not depend on the parameter as long as the orientation of S is preserved (Exercise 5). Thus the integral of ω over an oriented C^1 curve S of M is defined. When we reverse the orientation, traversing S in the opposite sense, it changes the sign of the integral. We write $\int_{-S} \omega = -\int_S \omega$.

More generally, let $\tilde{S}$ be an oriented continuous and piecewise differentiable curve, that is, we consider $\tilde{S}$ to be a union of curves $S_1, S_2, \ldots, S_r$ such that each S_i is C^1 and the terminal point of S_i is the initial point of S_{i+1} (terminal and initial point make sense since we are dealing with oriented curves). Then we define the integral over $\tilde{S}$ by

$$\int_{\tilde{S}} \omega = \sum_{i=1}^{r} \int_{S_i} \omega,$$

thus extending to this case the definition of line integral on a manifold. This definition reduces to the usual one when $M = \boldsymbol{R}^n$. In fact we could have used that as a starting point by subdividing the curve $\tilde{S}$ on an arbitrary manifold into a finite union of C^1 curves S_i, each in a single coordinate neighborhood and evaluated the integral over each S_i in local coordinates, that is in $\boldsymbol{R}^n$.

(5.6) Example Consider the special case $\omega = df$, where f is a C^∞ function on M. (This implies that $d\omega = 0$.) In this case the value of the line integral along the piecewise differentiable curve $\tilde{S}$ from p to q is given by

$$\int_{\tilde{S}} df = f(q) - f(p).$$

In particular, it is independent of the path chosen. The verification is Exercise 1 at the end of the section. Note that if p is held fixed, then $f(q)$ is

given at each q by adding $f(p)$ to the value of the line integral along any piecewise C^1 curve from p to q. Thus f is determined to within an additive constant by the line integral (assuming M connected).

The fact that a (line) integral of a one-form ω over an oriented piecewise differentiable curve $\tilde{S}$ has been defined enables us to state Stokes's theorem for a polygonal region Q of $\boldsymbol{R}^2$ bounded, as it is, by an oriented piecewise linear (simple closed) curve $\tilde{S} = \partial Q$. We carry this out for the unit square.

(5.7) Theorem *Let ω be a C^1 one-form defined on $Q = \{(x, y)\colon 0 \le x \le 1, 0 \le y \le 1\}$ and let $\tilde{S}$ be the boundary of Q traversed in the counterclockwise sense. Then $\int_Q d\omega = \int_{\tilde{S}} \omega$.*

Proof Let $\omega = a\,dx + b\,dy$, where a, b vanish outside Q and are C^1 functions on Q. Then $d\omega = ((\partial b/\partial x) - (\partial a/\partial y))\,dx \wedge dy$ on Q and by Remark 2.3,

$$\int_Q d\omega = \int_0^1 \int_0^1 \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right) dx\,dy$$

$$= \int_0^1 [b(1, y) - b(0, y)]\,dy - \int_0^1 [a(x, 1) - a(x, 0)]\,dx$$

The orientation is that given by the standard coordinate system in $\boldsymbol{R}^2$. On the other hand the integral over the boundary is

$$\int_{\tilde{S}} \omega = \sum_{i=1}^{4} \int_{S_i} a\,dx + b\,dy$$

$$= \int_0^1 a(x, 0)\,dx + \int_0^1 b(1, y)\,dy + \int_1^0 a(x, 1)\,dx + \int_1^0 b(0, y)\,dy.$$

This is because $dy = 0$ on the vertical sides and $dx = 0$ on the horizontal ones. (See Fig. VI.9.) Comparing the values of the integrals shows that the theorem is true. ∎

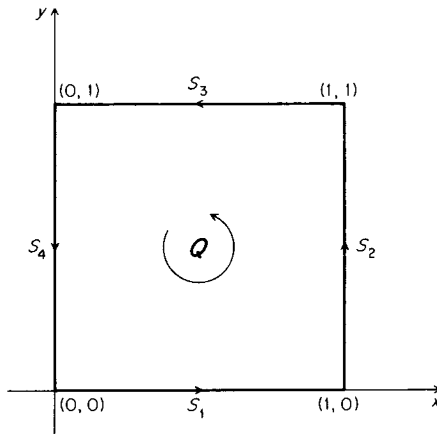


Figure VI.9

We remark that this essentially mimics our earlier proof, and from the procedure used it is clear that we could state and prove this theorem for rectangles of any dimension and even for triangles, tetrahedra, and other simplices although in these cases, as we mentioned, some machinery is necessary to describe proper orientations on the faces.

Exercises

1. Prove the formula of Example 5.6 for $\int_S df$ by first proving it for curves of class C^1 which lie in a coordinate neighborhood.
2. Evaluate the line integral of $\omega = x^2 y\, dx + x\, dy$ on $M = \boldsymbol{R}^2$ along the radial path from $(0, 0)$ to $(1, 1)$ and along the path consisting of the segments $(0, 0)$ to $(1, 0)$ and $(1, 0)$ to $(1, 1)$. Determine whether this integral is independent of the path.
3. Show that all line integrals of $\omega = P\, dx + Q\, dy + R\, dz$ in $\boldsymbol{R}^3$ are independent of the path only if the value of the integral over any closed (piecewise C^1) path is zero. Use this and Stokes's theorem to obtain a condition on P, Q, R which is sufficient to show independence of the path. [Assume ω is defined on all of $\boldsymbol{R}^3$.]
4. In Example 5.3 let M be the unit ball and ∂M the unit sphere. If $P = x^2$, $Q = y^2$, and $R = z^2$, compute both sides of the equation giving Stokes's theorem.
5. Suppose that S is an (oriented) C^1 curve on a manifold M given parametrically by a mapping $F: t \mapsto F(t)$, $a \le t \le b$ of $[a, b]$ into M. Suppose $t = f(s)$, $c \le s \le d$ is a change of parameter on S. Show that the value of the line integral over S of any one-form ω is unchanged if $f'(s) > 0$, that is, if the orientation of $G = F \circ f$ is the same.
6. Prove Stokes's theorem for a triangle in $\boldsymbol{R}^2$ and a cube in $\boldsymbol{R}^3$.

6 Homotopy of Mappings. The Fundamental Group

One of the most basic ideas used in the study of mappings from one space to another is that of homotopy. Two mappings are said to be homotopic if one can be "deformed" into the other through a one-parameter family of mappings between the same spaces. Sometimes further conditions are imposed on the family of mappings as we shall see. The basic definition can be stated as follows.

(6.1) Definition Let F, G be continuous mappings from a topological space X to a topological space Y and let $I = [0, 1]$, the unit interval. Then F

is *homotopic* to G if there is a continuous mapping (the *homotopy*)

$$H: X \times I \to Y$$

which satisfies the conditions: $F(x) = H(x, 0)$ and $G(x) = H(x, 1)$ for all $x \in X$. If X and Y are manifolds and F, G are C^∞, we define a C^∞ or *smooth homotopy* by requiring that H be C^∞ in addition to the conditions above.

We remark that $H_t(x) = H(x, t)$ does indeed define a one-parameter family of mappings $H_t: X \to Y$, $0 \le t \le 1$, with $F = H_0$ and $G = H_1$. The formulation of the definition emphasizes the simultaneous continuity in both variables t and x.

Some brief comments on the C^∞ case: If $\partial X = \varnothing$, then $X \times I$ is a regular domain of $X \times \boldsymbol{R}$ and is a manifold with boundary. Indeed, $\partial(X \times I) = X \times \{0\} \cup X \times \{1\}$, so C^∞ is perfectly well defined. If $\partial X \neq \varnothing$, then $X \times I$ is not a manifold with boundary [consider $X = \bar{B}_1^2(0)$, the closed unit disk, for example]. However, it is a reasonably nice domain of $X \times \boldsymbol{R}$ which is a manifold (with nonempty boundary), so only minor technical problems arise. We remark however, that when both X and Y have non-empty boundaries, there are cases in which it is natural to require that $H_t(\partial X) \subset \partial Y$ for $0 \le t \le 1$, which is closely related to the generalization below.

When the class of continuous maps from a space X to a space Y is considered in its entirety, then homotopy of maps forms an equivalence relation, and for many purposes it is the equivalence class of the map that is important and not the particular representative. We shall illustrate this in great detail in a special case, namely $X = I$, the unit interval. Before doing this we mention a useful generalization of our definition: Suppose (X, A) and (Y, B) are pairs consisting of spaces X and Y and closed subspaces $A \subset X$ and $B \subset Y$. Consider $F, G: X \to Y$ continuous maps such that $F(A) \subset B$ and $G(A) \subset B$; F and G map the pair (X, A) into the pair (Y, B) continuously. We say that F and G are *relatively homotopic* if there exists a continuous map $H: X \times I \to Y$ such that $H(A \times I) \subset B$, $H(x, 0) \equiv F(x)$, and $H(x, 1) \equiv G(x)$. We have added to Definition 6.1 the requirement that $H_t(A) \subset B$ for $0 \le t \le 1$. When $A = \varnothing = B$, the definition reduces to the original one. We will write $F \sim G$ to indicate that F and G are (relatively) homotopic; we justify this notation as follows.

(6.2) Theorem *Relative homotopy is an equivalence relation on the continuous maps of (X, A) into (Y, B) for any topological spaces X and Y and closed subspaces A and B, respectively.*

Proof The relation is reflexive since $H(x, t) \equiv F(x)$ is a homotopy of $F(x)$ with $F(x)$. It is symmetric as well; given a homotopy $H(x, t)$ of F to G,

then $\tilde{H}(x, t) = H(x, 1 - t)$ is a homotopy of G to F. Finally, suppose $F_1 \sim F_2$ and $F_2 \sim F_3$ by homotopies H_1 and H_2, respectively. Then we define $H(x, t)$ a homotopy of F_1 and F_3 by

$$H(x, t) = \begin{cases} H_1(x, 2t), & 0 \le t \le \frac{1}{2}, \\ H_2(x, 2t - 1), & \frac{1}{2} \le t \le 1. \end{cases}$$

It is easily verified that $H(x, t)$ is continuous, and that all these maps take A into B for every t between 0 and 1 inclusive. We leave the C^∞ case to the exercises. ∎

Homotopy of Paths and Loops. The Fundamental Group

As a first application of the concept of homotopy we will consider homotopy classes of continuous maps of the unit interval $I = [0, 1]$ into a manifold M. A map $f: I \to M$ of this type is called a *path*, $f(0)$ its *initial* point, and $f(1)$ its *terminal* point. We shall consider homotopy classes of paths under the additional restriction that the homotopy keep initial and terminal points fixed, that is, $H(t, 0)$ and $H(t, 1)$ are constant functions.† This is exactly relative homotopy for $(I, \{0, 1\})$ and $(X, \{b, d\})$, $b = f(0)$, $d = f(1)$. Given a manifold M, fix a basepoint b on M and consider the paths with b as initial point. If b is also the terminal point, then the path is called a *loop*; thus a loop is a continuous map $f: I \to M$ such that $f(0) = b = f(1)$. We denote its homotopy class by $[f]$, meaning always relative homotopy. Among these classes is that of the *constant loop* $e_b(s) \equiv b$, $0 \le s \le 1$. If this is the only homotopy class and M is connected, then we say M is *simply connected*; this means that every loop at b can be deformed over M to the constant loop. It is rather easy to see (Exercise 1) that this property does not depend on the choice of b and is equivalent to the statement that any closed curve (continuous image of S^1) may be continuously deformed to a point on M.

Paths, loops, and their homotopy classes are very useful in the study of spaces from the point of view of algebraic topology, for an important objective is to assign algebraic objects, such as groups, to spaces in such a way that they depend only on the topology of the space, that is, are invariant under homeomorphism, and thus "measure" topological features. We shall illustrate this process in this chapter for the case where our spaces are manifolds. The restriction to manifolds is not essential but is for convenience only.

If M is a connected manifold and f, g are paths on M with the terminal point $f(1)$ coinciding with the initial point $g(0)$, we may clearly combine these to a single path h after readjusting the parametrization; in fact,

$$h(s) = \begin{cases} f(2s), & 0 \le s \le \frac{1}{2}, \\ g(2s - 1), & \frac{1}{2} \le s \le 1, \end{cases}$$

† In our notation, here the *first* variable of H is the deformation parameter t.

is obviously a continuous map $h: I \to M$ traversing the image of f followed by that of g. We shall call this the *product* of f and g, denoted $f * g$. This product has the following properties with respect to (relative) homotopy:

(i) $f * (g * h) \sim (f * g) * h$.

(ii) Let $f(1) = b = g(0)$ and suppose $f = e_b$. Then $e_b * g \sim g$. Similarly, if $g = e_b$, then $f * e_b \sim f$.

(iii) If $f_1 \sim f_2$ and $g_1 \sim g_2$, then $f_1 * g_1 \sim f_2 * g_2$.

(iv) If $g(s) = f(1 - s)$, and $a = f(0)$, $b = f(1)$, then $f * g \sim e_a$ and $g * f \sim e_b$.

(v) If $F: M \to N$ is continuous and $f' = F \circ f$, $g' = F \circ g$, then $(f * g)' = f' * g'$.

The verification of most of these properties is left as an exercise. In each case a homotopy $H(t, s)$ having the given properties must be constructed; as a sample we verify (ii). By definition $e_b * g(s) = b$ for $[0, \frac{1}{2}]$ and $e_b * g(s) = g(2s - 1)$ for $[\frac{1}{2}, 1]$. We define $H(s, t)$ in the following way:

$$H(t, s) = \begin{cases} b, & 0 \le s \le \frac{1}{2}(1 - t) \quad \text{and} \quad 0 \le t \le 1, \\ g\left(\dfrac{2s - 1 + t}{1 + t}\right), & \frac{1}{2}(1 - t) \le s \le 1. \end{cases}$$

It is useful to see how this map $H: I \times I \to M$ maps various portions of the unit square in Fig. VI.10. The shaded portion is mapped onto $b = g(0)$ and each horizontal segment in the unshaded part, as for example the dotted line, is mapped onto the image of g with the parametrization modified proportionately. For property (iv) for $f * g$ we have a diagram as in Fig. VI.11 with the base mapping according to $f * g$, the top and two sides mapping onto a and the top and two sides of each inner square going onto $f(\frac{1}{2}(1 - t)) = g(\frac{1}{2}(1 + t))$. In verifying properties (i), (ii), and (iv) (Exercise 7) such diagrams are useful.

We are now ready to give an example of a group "assigned" to a manifold M, the *fundamental group* of M (at the basepoint b). It is an important

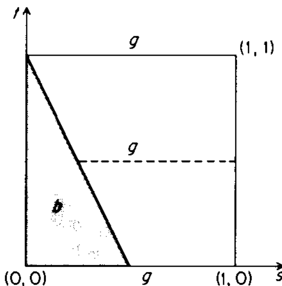


Figure VI.10

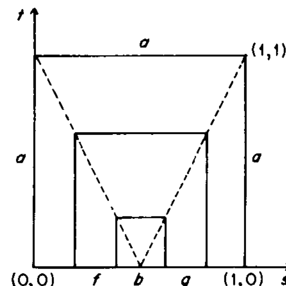


Figure VI.11

algebraic invariant of a topological space or manifold and is often called the *Poincaré group* after one of the founders of algebraic topology.

(6.3) Theorem *Let $\pi_1(M, b)$ denote the homotopy classes of all loops at $b \in M$. Then $\pi_1(M, b)$ is a group with product $[f][g] = [f * g]$. If $F: M \to N$ is continuous, then F determines a homomorphism $F_*: \pi_1(M, b) \to \pi_1(N, F(b))$ by $F_*[f] = [F \circ f]$. If G is homotopic to F relative to the pairs (M, b) and $(N, F(b))$, then $F_* = G_*$. Finally, when F is the identity mapping on M, F_* is the identity isomorphism, and $(F \circ G)_* = F_* \circ G_*$ for compositions of continuous mappings.*

Proof The product is well defined since (iii) above assures us that $[f * g]$ is independent of the representatives f and g chosen from $[f]$ and $[g]$. By (i) it is associative; and (ii) and (iv) give the existence of an identity $[e_b]$ and inverse. Thus $\pi_1(M, b)$ is a group. It follows from (v) that $F: M \to N$ induces a homomorphism F_*, and the last statement of the theorem is immediate from the definitions. Finally if $H: M \times I \to N$ is a homotopy of F and G, then $\tilde{H}(t, s) = H(f(s), t)$ is a homotopy of the loops $F \circ f$ and $G \circ f$; hence $F_*[f] = G_*[f]$. ∎

We have some immediate corollaries, the first of which spells out the meaning of the statement that the fundamental group is a topological "invariant."

(6.4) Corollary *If M_1 and M_2 are homeomorphic by $F: M_1 \to M_2$ and $F(b_1) = b_2$, then F_* is an isomorphism of the corresponding fundamental groups $\pi_1(M_1, b_1) \cong \pi_2(M_2, b_2)$.*

Proof If $F: M_1 \to M_2$ is the homeomorphism and $G: M_2 \to M_1$ its inverse, then F_* and G_* are isomorphisms, since $F_* \circ G_*$ and $G_* \circ F_*$ are the identity isomorphisms by the last statement of the theorem. ∎

If the identity map of M to M is homotopic to the constant map of M onto one of its points b, then M is said to be *contractible* (to b). For example, any open subset of $\boldsymbol{R}^n$ which is star-shaped with respect to a point b is contractible since $H(x, t) = (1 - t)x + tb$ is such a homotopy. For contractible spaces we have the following:

(6.5) Corollary *If M is contractible to b, then $\pi_1(M, b) = \{e\}$, the identity element alone. It follows that M is simply connected.*

Proof If f is a loop at b, then it is homotopic to the constant loop e_b by $\tilde{H}(t, s) = H(f(s), t)$. This shows that $\pi_1(M, b) = \{1\}$. To deduce simple connectedness from this is exactly Exercise 1. It is even simpler to prove it

directly from the definition using again the mapping H. Of course, there are simply connected spaces which are not contractible, the sphere S^n, $n > 1$, being the simplest example (see Corollary 7.14 below). ∎

An interesting application of these ideas arises when we consider line integrals along piecewise differentiable paths on M. Let ω be a one-form on M and suppose $p, q \in M$. If S_1, S_2 are two such paths of M from p to q it is natural to ask whether or not

$$\int_{S_1} \omega = \int_{S_2} \omega.$$

One knows that in general they are not equal, even in very simple cases (see Exercise 10). However, the standard theorems of advanced calculus on independence of path may be generalized to manifolds with essentially the same proofs. We shall state the results and sketch the proofs.

(6.6) Theorem *Let ω be a one-form on a manifold M such that $d\omega = 0$ everywhere, and let S_1, S_2 be homotopic piecewise differentiable paths from $p \in M$ to $q \in M$. Then*

$$\int_{S_1} \omega = \int_{S_2} \omega.$$

Proof (*in outline*) If S_1 and S_2 are C^1 curves homotopic by a differentiable mapping H of $I \times I$ into M, then this result is a straightforward application of Theorem 5.7 (Stokes's theorem for the unit square. In the general case the (continuous) homotopy H of the piecewise differentiable curves must be altered as follows. First $I \times I$ is subdivided by vertical and horizontal lines (Fig. VI.12) so that it is differentiable on each boundary segment and so that H carries each subrectangle Q_{ij} into a single coordinate neighborhood U. Then the techniques of Theorem V.4.8 are used to alter H successively to a homotopy $\tilde{H}$ which is *differentiable* on each Q_{ij}. From this point the proof follows the usual one of advanced calculus. The argument is as follows:

The new homotopy $\tilde{H}$ maps the edges of the square $Q = I \times I$ into the paths S_1, q, $-S_2$, p, respectively, as we go around ∂Q counterclockwise. The images of the left and right vertical edges are the constant paths p and q. (See Fig.VI.12.) Since the line integral of ω over a constant path is zero, we have

$$\int_{\partial Q} \tilde{H}^*\omega = \int_{S_1} \omega + \int_{-S_2} \omega = \int_{S_1} \omega - \int_{S_2} \omega.$$

On the other hand, it is easy to check that if we denote the oriented squares of the subdivision by Q_{ij}, then line integrals over the same path in opposite directions cancel out, and we have

$$\int_{\partial Q} \tilde{H}^*\omega = \sum_{i,j} \int_{\partial Q_{ij}} \tilde{H}^*\omega.$$

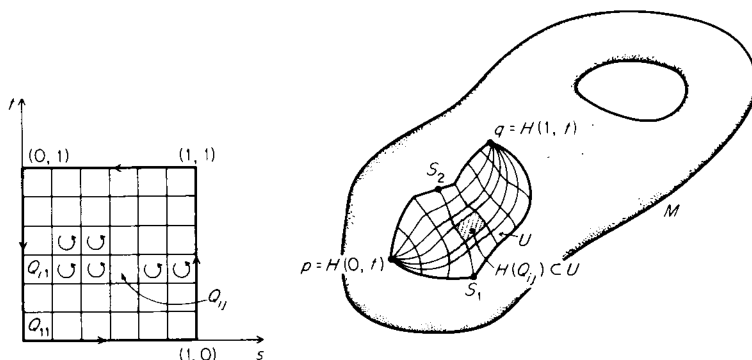


Figure VI.12

By Theorem 5.7 and the remarks preceding it,

$$\int_{\partial Q_{ij}} \tilde{H}^*\omega = \int_{Q_{ij}} d\tilde{H}^*\omega.$$

Since $d\tilde{H}^*\omega = \tilde{H}^*\, d\omega = 0$, we see that $\int_{S_1} \omega - \int_{S_2} \omega = 0$, which was to be proved. ∎

Just as in the case of domains of $\boldsymbol{R}^n$, this theorem has the following corollary:

(6.7) Corollary *Let ω be a C^∞ one-form on a simply connected manifold M and suppose that $d\omega = 0$ everywhere. Then there is a C^∞ function f on M such that $\omega = df$. If f and g are two such functions, then $f - g$ is constant.*

Proof We choose a fixed basepoint $b \in M$ and define f at any $p \in M$ by choosing a piecewise differentiable curve S from b to p and setting $f(p) = \int_S \omega$. Theorem 6.6 assures us that this defines a function on M. The remainder of the proof deals with purely local properties; we must show that f is a C^∞ function with the property that $df = \omega$. If we show the latter fact, it will follow that f is C^∞ because we have assumed ω to be C^∞. Changing the basepoint changes f by an additive constant—the integral of ω along the path between the old and new basepoints—hence does not change df at all; therefore it is enough to show that $df = \omega$ at the basepoint. Let U, φ be a coordinate neighborhood of the basepoint b. We suppose that $x^1, \ldots, x^n$ are the local coordinates and that $\varphi(b) = (0, \ldots, 0)$ and $\varphi(U) = B_1^n(0)$ and we let $f(x^1, \ldots, x^n)$ denote the expression for f in local coordinates. Then denoting ω in local coordinates by $\omega = \alpha_1(x)\, dx^1 + \cdots + \alpha_n(x)\, dx^n$, we have, by definition

$$f(x) = \int_C \alpha_1(x)\, dx^1 + \cdots + \alpha_n(x)\, dx^n,$$

the line integral along any path C from $(0, \ldots, 0)$ to $(x^1, \ldots, x^n)$. We must

show that $\partial f/\partial x^j = \alpha_j$, $j = 1, \ldots, n$, at $x = (0, \ldots, 0)$. However, this is immediate from the definitions:

$$\left(\frac{\partial f}{\partial x^j}\right)_0 = \lim_{h \to 0} \frac{1}{h}(f(0, \ldots, h, \ldots, 0) - f(0, \ldots, 0))$$
$$= \lim_{h \to 0} \frac{1}{h} \int_0^h \alpha_j(0, \ldots, x^j \ldots 0)\, dx^j$$
$$= \alpha_j(0, \ldots, 0).$$

This completes the proof, except for the last statement, which is obvious: $d(f - g) = \omega - \omega = 0$ so that $f - g$ = constant on the (connected) manifold M. ∎

We remark that in terms of the expression of ω in local coordinates, $d\omega = 0$ is equivalent to $(\partial\alpha_i/\partial x^j) - (\partial\alpha_j/\partial x^i) = 0$, $1 \leq i, j \leq n$. For one-forms on $\boldsymbol{R}^n$, this is usually stated by saying that the *curl* of the vector field $\alpha_1(\partial/\partial x^1) + \cdots + \alpha_n(\partial/\partial x^n)$ associated to ω vanishes (see Apostol [1]).

The concept of fundamental group and the techniques of this section are intimately related to the notion of covering manifold and of properly discontinuous group action on a manifold and will allow us to complete the study of these phenomena, begun at the end of Chapter III. We will do this in the last section of this chapter.

Exercises

1. Prove that a necessary and sufficient condition that every closed curve in a connected space M be continuously deformable to a point is that $\pi_1(M, b) = 1$ for some $b \in M$.
2. Let a, b be points of a connected manifold M and show that $\pi_1(M, a)$ and $\pi_1(M, b)$ are isomorphic.
3. Show that $\pi_1(M \times N, (a, b))$ is naturally isomorphic to $\pi_1(M, a) \times \pi_1(N, b)$.
4. Show that $\pi_1(S^1) \cong \boldsymbol{Z}$ and $\pi_1(S^2) = \{1\}$. Use this to show that S^2 and T^2 are not homeomorphic. [To show $\pi_1(S^2)$ is trivial one must show as a first step that any loop at N, the north pole, is homotopic to one contained in the punctured sphere $S^2 - \{x\}$ (where x is a point distinct from N).]
5. Show that if G is a connected Lie group, then $\pi_1(G)$ is Abelian. Use e as basepoint.
6. If $F, G\colon M \to N$ are continuous maps of C^∞ manifolds, F is homotopic to G, and $F(p) = q = G(p)$; then $F_*(\pi_1(M, p))$ and $G_*(\pi_1(M, p))$ are conjugate subgroups of $\pi_1(N, q)$. [*Note:* We do *not* assume that the homotopy H of F and G is constant on p.]

7. Verify properties (i)–(v) used in the proof of Theorem 6.3.
8. Prove that if we use only piecewise differentiable paths and piecewise differentiable homotopies to define $\pi_1(M, b)$, the group obtained is the same as when continuous paths and homotopies are used.
9. Show that C^∞ homotopy of C^∞ mappings of M to N, manifolds (without boundary), is an equivalence relation on the set of such mappings.
10. Show that on $\boldsymbol{R}^2$ the following integrals depend on the path chosen:
    (a) the line integral of $\omega = y\,dx + dy$ from $p = (0, 0)$ to $q = (1, 1)$,
    (b) the line integral of $\omega = (-y/r^2)\,dx + (x/r^2)\,dy$, $r^2 = x^2 + y^2$ [in the latter case, we exclude $(0, 0)$].
    In either case is $d\omega = 0$?

7 Some Applications of Differential Forms. The de Rham Groups

It is our purpose in this section to obtain a few results about manifolds which are traditionally in the domain of algebraic topology. We do not assume a knowledge of this subject, but we have mentioned briefly in the Exercises to Section V.8 the following definitions, with some consequences which follow from the results of that section.

(7.1) Definition A k-form ω on a manifold M (with possibly nonempty boundary) is said to be *closed* if $d\omega = 0$ everywhere and is said to be *exact* if there is a $(k-1)$-form η such that $d\eta = \omega$.

We recall some facts about the operator d and apply them here. Let $Z^k(M)$ denote the closed k-forms on M; since $Z^k(M)$ is the kernel of the homomorphism $d\colon \bigwedge^k(M) \to \bigwedge^{k+1}(M)$ it is a linear subspace of $\bigwedge^k(M)$. Similarly the exact k-forms $B^k(M)$ are the image of $d\colon \bigwedge^{k-1}(M) \to \bigwedge^k(M)$ and thus a linear subspace. Moreover $d^2 = 0$ implies that $B^k(M) \subset Z^k(M)$ which allows us to form the quotient $H^k(M)$.

(7.2) Definition The quotient space $H^k(M) = Z^k(M)/B^k(M)$ is called the *de Rham group of dimension* k of M. If $n = \dim M$, we denote by $H^*(M)$ the direct sum

$$H^*(M) = H^0(M) \oplus \cdots \oplus H^n(M).$$

Note that $H^*(M) = Z(M)/B(M)$, where $Z(M)$ and $B(M)$ are the kernel and image of $d\colon \bigwedge(M) \to \bigwedge(M)$, respectively (and the direct sums of the $Z^k(M)$ and $B^k(M)$, $k = 0, 1, \ldots, n$). Although called de Rham *groups*, $H^k(M)$, $k = 0, \ldots, n = \dim M$, are actually vector spaces over $\boldsymbol{R}$ and, in fact, it is easy to verify that $H^*(M)$ is an *algebra* with the multiplication being that naturally induced by the exterior product of differential forms. This follows

directly from the basic property of d which asserts that when $\varphi \in \bigwedge^r(M)$, $\psi \in \bigwedge^s(M)$, then

$$d(\varphi \wedge \psi) = d\varphi \wedge \psi + (-1)^r \varphi \wedge d\psi$$

(from which it follows that $Z(M)$ is an algebra containing $B(M)$ as an ideal). The great importance of the de Rham groups and the de Rham algebra $H^*(M)$ stems from de Rham's theorem, which we state in somewhat abbreviated and imprecise form, for its general interest.

(7.3) Theorem *There is a natural isomorphism of $H^*(M)$ and the cohomology ring of M under which $H^k(M)$ corresponds to the* kth *cohomology group.*

Since we do not assume any knowledge of algebraic topology the cohomology groups will not be defined here, nor will any proof of this important theorem be attempted. The reader is referred to Warner [1] and the references found there. We do remark, however, that among the consequences are the facts that whenever M is compact the dimension of $H^*(M)$ is finite and that in any case $H^*(M)$ together with its structure as an algebra are topologically invariant, that is, if M_1 and M_2 are homeomorphic, then $H^*(M_1)$ and $H^*(M_2)$ are isomorphic as algebras. Finally we mention that the duality which appears in algebraic topology between homology and cohomology groups of a space extends to a duality of homology groups and de Rham groups via integration and Stokes's theorem—a further motivation for the earlier sections of this chapter. Of particular interest to us is the fact that we can use these de Rham groups, without using algebraic topology, to obtain interesting results about manifolds. Samples are given below and in the following section.

It is a basic property of differential forms that a C^∞ mapping $F: M_1 \to M_2$ defines a corresponding homomorphism $F^*: \bigwedge(M_2) \to \bigwedge(M_1)$. Since $F^*d = dF^*$, it follows that $F^*(Z^k(M_2)) \subset Z^k(M_1)$ and $F^*(B^k(M_2)) \subset B^k(M_1)$. Therefore F^* induces a homomorphism, which we also denote by F^*, of $H^k(M_2)$ into $H^k(M_1)$. Since F^* is an algebra homomorphism on forms, $F^*: H^*(M_2) \to H^*(M_1)$ is also an algebra homomorphism. In summary, with the above notation we have the following lemma:

(7.4) Lemma *A C^∞ mapping $F: M_1 \to M_2$ induces an algebra homomorphism $F^*: H^*(M_2) \to H^*(M_1)$ which carries $H^k(M_2)$ (linearly) into $H^k(M_1)$ for all k. If F is the identity mapping on M, then $F^*: H^*(M) \to H^*(M)$ is the identity isomorphism. Under composition of mappings we have $(G \circ F)^* = F^* \circ G^*$.*

Using this lemma, we can obtain a weak version of the invariance of $H^*(M)$ under homeomorphism mentioned above.

(7.5) Corollary *If M_1 and M_2 are diffeomorphic manifolds, then $H^*(M_1)$ and $(H^*(M_2)$ are isomorphic as algebras.*

Proof Let $F: M_1 \to M_2$ be a diffeomorphism and F^{-1} its inverse. Then $F^{-1*} \circ F^* = (F \circ F^{-1})^*$ and $F^* \circ F^{-1*} = (F^{-1} \circ F)^*$ are both the identity isomorphism, hence F^* is an isomorphism with inverse F^{-1*}. ∎

Although the groups $H^k(M)$ are difficult to compute using only the tools which we have available, which do not include algebraic topology, we can obtain information in special cases—information which we can then use in some applications.

(7.6) Theorem *Let M be a C^∞ manifold with a finite number r of components. Then $H^0(M) = V^r$, a vector space over $\boldsymbol{R}$ of dimension r.*

Proof $\bigwedge^0(M)$ consists of C^∞-functions on M and $Z^0(M)$ of those functions f for which $df = 0$. There are no forms of dimension less than zero so $B^0(M) = \{0\}$ and $H^0(M) = Z^0(M)$. We have seen previously that $df = 0$ if and only if f is constant on each component $M_1, \ldots, M_r$. Thus $H^0(M) \cong \{(a_1, \ldots, a_r)\colon a_i \in R\}$, where $(a_1, \ldots, a_r)$ corresponds to the function taking the constant value a_i on M_i, $i = 1, \ldots, r$. ∎

(7.7) Remark For $M = \{p\}$, a point, $\bigwedge^k(M) = \{0\}$ if $k \neq 0$ and $\bigwedge^0(M) \cong \boldsymbol{R}$. We define $d\omega = 0$ for all forms ω. It follows that $H^0(\{p\}) \cong \boldsymbol{R}$, $\{p\}$ being a zero-dimensional manifold; this determines the de Rham groups of a point space—since $\bigwedge^k(\{p\}) = 0$, $H^k(\{p\}) = 0$ for $k > 0$.

As an immediate consequence of Corollary 6.7 and Theorem 7.6, we have the following theorem:

(7.8) Theorem *If M is a simply connected manifold, then $H^0(M) \cong \boldsymbol{R}$ and $H^1(M) = \{0\}$.*

Proof Suppose ω is a closed one-form on M, that is, $d\omega = 0$. Then there exists a function f on M such that $df = \omega$, thus ω is exact. Since every closed one-form is exact, $H^1(M) = \{0\}$. $H^0(M) \cong \boldsymbol{R}$ since M is connected. ∎

In addition to this information concerning $H^0(M)$ and $H^1(M)$ we may also prove the following statement concerning the highest-dimensional de Rham group $H^n(M)$, $n = \dim M$.

(7.9) Theorem *Let M be a compact orientable manifold of dimension n with $\partial M = \varnothing$. Then there is a natural homomorphism of $H^n(M)$ onto R. In particular, $H^n(M) \neq \{0\}$*

Proof Since dim $M = n$, every n-form on M is closed; hence $Z^n(M) = \bigwedge^n(M)$. Define a homomorphism $h\colon Z^n(M) \to \boldsymbol{R}$ by $h(\Omega) = \int_M \Omega$. If $\Omega \in B(M)$, then $\Omega = d\omega$ for some $\omega \in \bigwedge^{n-1}(M)$. Since $\partial M = \varnothing$, it follows from Stokes's theorem that

$$h(\Omega) = h(d\omega) = \int_M d\omega = \int_{\partial M} \omega = 0.$$

Therefore h determines a homomorphism $h'\colon H^n(M) \to \boldsymbol{R}$. Let Ω_0 be a volume element on M. Then $h(\Omega_0) = \int_M \Omega_0 \neq 0$ by Theorem 2.2. Thus, by suitable choice of a constant c, we can make $h(c\Omega_0) = c \int_M \Omega_0$ take any real number as value. Hence h and h' are onto. ∎

The Homotopy Operator

In order to obtain some further results concerning de Rham groups we will introduce a special operator $\mathscr{I}$, the *homotopy operator*. Let $A \subset \boldsymbol{R}^n$ be such that $\bar{A}$ is the closure of the interior of A. We have in mind regular domains, cubes, simplices, or their interiors. Note that $I \times \bar{A}$ is the closure of an open set, its own interior, in $\boldsymbol{R} \times \boldsymbol{R}^n = \boldsymbol{R}^{n+1}$. When A is not open, a C^∞ k-form ω on A is the restriction to A of a k-form $\tilde{\omega}$ on an open set $U \supset A$—by definition of differentiability of functions (in this instance its components) on A. Our restrictions on A ensure that all derivatives of any C^∞ function f on A are defined at every $p \in A$ independently of the open set U and extension $\tilde{f}$ which may be needed to define them at boundary points. This is a consequence of the continuity of all derivatives of $\tilde{f}$ on U and of the fact that every $p \in A$ is either an interior point—where the derivatives are already defined without any $\tilde{f}$—or the limit of interior points. It follows that for a C^∞ form ω on A, $d\omega$ is defined, even at boundary points.

(7.10) Definition The *homotopy operator* $\mathscr{I}$ is defined to be an $\boldsymbol{R}$-linear operator from $\bigwedge^{k+1}(I \times A) \to \bigwedge^k(A)$. On monomials $\mathscr{I}$ is defined as follows: If $\omega = \alpha(t, x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k+1}}$, we set $\mathscr{I}\omega = 0$; and if $\omega = \alpha(t, x)\, dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, we define $\mathscr{I}\omega$ by

$$\mathscr{I}\omega = \left(\int_0^1 \alpha(t, x)\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

Having been thus defined for monomials, we extend $\mathscr{I}$ to be $\boldsymbol{R}$-linear on $\bigwedge^{k+1}(I \times A)$ with values in $\bigwedge^k(A)$.

We will denote by $i_t\colon A \to I \times A$ the natural injection $i_t(x) = (t, x)$ and then ω_t will denote $i_t^*\omega$; in particular $\omega_0 = i_0^*\omega$ and $\omega_1 = i_1^*\omega$. With these definitions and notations we find that $\mathscr{I}$ has the following basic properties.

(7.11) Lemma *The homotopy operator* $\mathscr{I}: \bigwedge^{k+1}(I \times A) \to \bigwedge^k(A)$ *in addition to being* $\boldsymbol{R}$*-linear has the following properties:*

(i) *it commutes with* C^∞ *functions which are independent of* t*;*

(ii) *for all* $\omega \in \bigwedge^{k+1}(I \times A)$ *it satisfies the relation*

$$\mathscr{I}\, d\omega + d\mathscr{I}\omega = \omega_1 - \omega_0 .$$

Proof If f is independent of t, we may consider it both as a function on $I \times A$ and on A. Since it is independent of t it can be moved through the integral sign in the definition of $\mathscr{I}$, thus $\mathscr{I} f\omega = f\mathscr{I}\omega$.

For the second property we must verify the equation directly; it is enough to do so for monomials since d, $\mathscr{I}$, i_0^*, and i_1^* are all $\boldsymbol{R}$-linear. First we consider the case where ω does not involve dt, in other words $\omega = \alpha(t, x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k+1}}$. Then $\mathscr{I}\omega = 0$ so that $d\mathscr{I}\omega = 0$ and $\mathscr{I}\, d\omega$ is given by

$$\mathscr{I}\, d\omega = \left(\int_0^1 \frac{\partial \alpha}{\partial t}\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_{k+1}} = (\alpha(1, x) - \alpha(0, x))\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k+1}}.$$

But the right side is then exactly $i_1^*\omega - i_0^*\omega = \omega_1 - \omega_0$, which thereby establishes the equality for this case.

Now suppose that $\omega = \alpha(t, x)\, dt \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}$. Computing $\mathscr{I}\, d\omega$, we see that

$$\mathscr{I}\, d\omega = -\sum_{j=1}^{n} \left(\int_0^1 \frac{\partial \alpha}{\partial x^j}\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

On the other hand using the Leibniz rule to differentiate under the integral sign (see Exercise 1.6), we may compute $d\mathscr{I}\omega$:

$$\begin{aligned} d\mathscr{I}\omega &= d\left(\int_0^1 \alpha(t, x)\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &= \sum_{j=1}^{n} \left(\int_0^1 \frac{\partial \alpha}{\partial x^j}\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \end{aligned}$$

Adding these expressions we see that $\mathscr{I}\, d\omega + d\mathscr{I}\omega = 0$. On the other hand since $i_1^*\, dt = 0 = i_0^*\, dt$, we have $0 = i_1^*\omega - i_0^*\omega = \omega_1 - \omega_0$. Thus in all cases (ii) holds. ∎

We continue to denote by A a subset of $\boldsymbol{R}^n$ which is either open or is the closure of an open set, and we use Lemma 7.11 to obtain *Poincaré's lemma.*

(7.12) Lemma *If A is star-shaped, then $H^k(A) = \{0\}$ for all $k \geq 1$. Hence $H^*(A)$ is isomorphic to the cohomology ring of a point. In particular, this holds for $A = \boldsymbol{R}^n$.*

Proof We recall that A is star-shaped if it contains a point 0 such that for any $p \in A$, the segment $\overline{0p}$ lies entirely in A. By suitable choice of coordinates we may suppose that 0 is the origin. We define $H: I \times A \to A$ as

$$H(t, x^1, \ldots, x^n) = (tx^1, \ldots, tx^n).$$

If ω is a k-form on A, then $H^*\omega$ is a k-form on $I \times A$. In the definition of $\mathscr{I}$ we defined $i_0: x \to (0, x)$ and $i_1: x \to (1, x)$; therefore $H \circ i_0: A \to \{0\}$ and $H \circ i_1: A \to A$ is the identity. For $k > 0$, $i_0^*(H^*\omega) = 0$ since $\bigwedge^k(\{0\}) = 0$ for a point space $\{0\}$. Thus,

$$d\mathscr{I}(H^*\omega) + \mathscr{I}\, d(H^*\omega) = i_1^*(H^*\omega),$$

so that if $d\omega = 0$, then $dH^*\omega = 0$ and

$$d\mathscr{I}H^*\omega = (H \circ i_1)^*\omega = \omega.$$

Therefore every closed k-form ω on A is exact if $k \geq 1$. If $k = 0$, then we may use the fact that A is connected to see that $H^0(A) \cong \boldsymbol{R}$. ∎

In fact, the homotopy operator $\mathscr{I}$ can be defined and used under more general hypotheses. We have supposed that A is a particular type of subset of $\boldsymbol{R}^n$, but it is possible to extend the definition to manifolds with or without boundary. As a sample we shall prove the following theorem.

(7.13) Theorem *Let M and N be C^∞ manifolds and assume $\partial M = \varnothing$. Suppose that F and G are C^∞ mappings of M into N which are C^∞ homotopic. Then the corresponding homomorphisms F^* and G^* of $H^*(N)$ into $H^*(M)$ are equal.*

Proof We shall use our previously defined operator $\mathscr{I}$ to construct a similar operator $\mathscr{I}: \bigwedge^{k+1}(I \times M) \to \bigwedge^k(M)$. First we note that M may be covered by a collection of coordinate neighborhoods, U_i, φ_i with $\varphi_i(U_i) = B_1^n(0)$, $n = \dim M$ and $i = 1, 2, \ldots,$ for which we have a subordinate C^∞ partition of unity $\{f_i\}$, supp $f_i \subset U_i$. Then any $(k+1)$-form ω on $I \times M$ can be written as a sum of forms with support in $I \times U_i$,

$$\omega = \sum \omega_i \qquad \omega_i = \begin{cases} f_i\omega & \text{on} \quad U_i, \\ 0 & \text{on} \quad M - \operatorname{supp} f_i. \end{cases}$$

This sum is, of course, finite on a neighborhood of each $p \in M$. We may consider f_i, or any functions on M, as being also functions on $I \times M$ which are independent of t. We define $\mathscr{I}$ to be additive so that $\mathscr{I}\omega = \sum \mathscr{I}\omega_i$, which leaves only the problem of defining $\mathscr{I}$ on forms with support in one of the neighborhoods $I \times U_i$.

When ω has support in a neighborhood $I \times U$, where U, φ is a coordinate neighborhood with $\varphi(U) = B_1^n(0)$, we proceed as follows. Let

$\tilde{\varphi}: I \times U \to I \times B_1^n(0)$ be defined by $\tilde{\varphi}(t, p) = (t, \varphi(p))$. Then define $\mathscr{I}\omega$ on $I \times U$, using our previous definition of $\mathscr{I}$ for $I \times B_1^n(0)$, by $\mathscr{I}\omega|_U = \tilde{\varphi}^*(\mathscr{I}(\tilde{\varphi}^{-1*}\omega))$, the $\mathscr{I}$ on the right side being the operator defined earlier, and further, let $\mathscr{I}\omega = 0$ on $M - U$. This defines a C^∞ k-form on M, the image of a $(k+1)$-form on $I \times M$. By Lemma 7.11 for this form ω we have the relation $\mathscr{I}d\omega + d\mathscr{I}\omega = \omega_1 - \omega_0$. Now since $\mathscr{I}d + d\mathscr{I}$ is an *additive* operator, then for an *arbitrary* $\omega \in \bigwedge^{k+1}(I \times M)$ we may apply the decomposition $\omega = \sum \omega_i$ to obtain

$$\begin{aligned}\mathscr{I}\, d\omega + d\mathscr{I}\omega &= \mathscr{I}\, d \sum \omega_i + d\mathscr{I} \sum \omega_i \\ &= \sum \mathscr{I}\, d\omega_i + \sum d\mathscr{I}\omega_i = \sum \left((\omega_i)_1 - (\omega_i)_0\right) \\ &= \omega_1 - \omega_0 .\end{aligned}$$

Finally, to complete the proof we let ω be any closed k-form on N and we must show that $G^*\omega - F^*\omega$ is exact. Now let $H: I \times M \to N$ be the homotopy connecting F and G. Then $F(p) = H(0, p) = H \circ i_0$ and $G(p) = H(1, p) = H \circ i_1$, where $i_t(p) = (t, p)$ as before. Since $dH^*\omega = H^*\, d\omega = 0$, we have then,

$$d\mathscr{I}H^*\omega = i_1^*H^*\omega - i_0^*H^*\omega = G^*\omega - F^*\omega,$$

as was to be shown. ∎

We say that a manifold M is C^∞ contractible if there is a C^∞ homotopy $H: I \times M \to M$ of the identity and a constant map, i.e., for all $q \in M$, $H(0, q) = q$ and $H(1, q) = p$, p fixed.

(7.14) Corollary *If M is C^∞ contractible, then it has the cohomology of a point space: $H^k(M) = \{0\}$ for $k > 0$ and $H^0(M) \cong \mathbf{R}$. In particular, a compact, orientable manifold is not C^∞ contractible.*

Proof This follows from the theorem with $M = N$ and Remark 7.7. For the last statement, if M were C^∞ contractible it would contradict Theorem 7.9. ∎

(7.15) Remark It is not difficult using theorems like Theorem V.4.8 to extend many of the results here and below to the case of *continuous* maps and homotopies.

Exercises

1. Let ω be a closed one-form on a compact manifold M. We define a mapping $F_\omega: \pi_1(M, b) \to \mathbf{R}$ by the following method. Let $f: [0, 1] \to M$ be a piecewise C^1 loop S at b and denote by $[f]$ the corresponding

element of $\pi_1(M, b)$. We define $F_\omega([f]) = \int_S \omega$. Show that F_ω is a homomorphism whose kernel contains the commutator subgroup of the fundamental group.

2. Compute $H^*(S^1)$ and $H^*(M)$. where M is a Möbius band.
3. Let M be a manifold, N a submanifold of M, and $R: M \to N$ a C^∞ mapping which leaves N pointwise fixed. Show that $R^*: H^*(N) \to H^*(M)$ is an injective homomorphism.
4. Let $M = M_1 \times M_2$ and let $P_i: M \to M_i$ be the natural projections. Show that $P_i^*: H^*(M_i) \to H^*(M)$ is an injective homomorphism.
5. Let $\bigwedge_c(M) \subset \bigwedge(M)$ denote the subalgebra of differential forms of compact support. Show that $H_c^*(M)$ may be defined analogously to $H^*(M)$ and is naturally isomorphic to a subalgebra. Prove that if M is connected and not compact, then $H_c^0(M) = \{0\}$.

8 Some Further Applications of de Rham Groups

In this section we use the information on de Rham groups accumulated in the previous section to obtain some very interesting facts which one customarily demonstrates by use of algebraic topology. In particular, it is proved here that there is no vector field on S^2 which does not vanish at some point—a fact to which we have frequently alluded. In a very beautiful little book, Milnor [2] has shown how many purely *topological* results can be obtained by differentiable methods. Using this idea and the results of Section 7, we give a demonstration of the Brouwer fixed point theorem, following the author's note (Boothby [1]).

We begin our proof of this last mentioned theorem by establishing a lemma. Let D^n denote $\bar{B}_1^n(0)$, the closed unit ball in $\mathbf{R}^n$. Then D^n is a manifold with boundary, $\partial D^n = S^{n-1}$. (For another proof, see Milnor [6].)

(8.1) Lemma *There is no C^∞ map $F: D^n \to \partial D^n$ which leaves ∂D^n pointwise fixed.*

Proof Suppose that there existed such a map F and let G denote the identity map of $\partial D^n \to D^n$. Then $F \circ G = I$, the identity map of $\partial D^n \to \partial D^n$. This implies that $G^* \circ F^* = (F \circ G)^*$ induces the identity isomorphism on $H^*(\partial D^n)$. Therefore the homomorphism $F^*: H^{n-1}(\partial D^n) \to H^{n-1}(D^n)$ must be injective, that is, $\ker F^* = \{0\}$. Since $H^{n-1}(D^n) = \{0\}$ by Poincaré's lemma, it follows $\ker F^* = H^{n-1}(\partial D^n)$ and therefore that $H^{n-1}(\partial D^n) = \{0\}$. However, because $\partial D^n = S^{n-1}$, an orientable and compact manifold without boundary, Theorem 7.9 shows that $H^{n-1}(\partial D^n) = H^{n-1}(S^{n-1}) \neq \{0\}$. This contradiction implies that no such map F exists. ∎

Using this lemma we may establish (as in Milnor [2]) a very well-known theorem of algebraic topology, the Brouwer fixed point theorem.

(8.2) Theorem (Brouwer) *Let X be a topological space homeomorphic to D^n. Then any continuous map $F: X \to X$ has a fixed point, that is, for each F there is at least one $x_0 \in X$ such that $F(x_0) = x_0$.*

Proof As a first step we note that it is enough to prove the theorem for D^n. Let $H: D^n \to X$ be a homeomorphism, and let $F: X \to X$ be any continuous mapping. If $H^{-1} \circ F \circ H: D^n \to D^n$ has a fixed point y_0, then $x_0 = H(y_0)$ is fixed by F.

Moreover, even in the case of D^n, it is enough to establish the property for C^∞ maps $F: D^n \to D^n$. To see this we suppose every such C^∞ map has a fixed point and assume there is some continuous map $G: D^n \to D^n$ which has no fixed point. Then $\|G(x) - x\|$ is bounded away from zero on the (compact) set D^n, and we may find an $\varepsilon > 0$ such that $\|G(x) - x\| > 3\varepsilon$. Using the Weierstrass approximation theorem (Lemma V.4.9), or the approximation theorem given in Exercise 2, we approximate G to within ε by a C^∞ mapping G_1; then $\|G(x) - G_1(x)\| < \varepsilon$ for all $x \in D^n$. However, since the values $G_1(x)$ are not necessarily in D^n for every $x \in D^n$, we replace G_1 by $F(x) = (1 + \varepsilon)^{-1} G_1(x)$. Clearly $F(x)$ is defined and C^∞ on D^n and $F(D^n) \subset D^n$. Since $\|G(x)\| \le 1$, it follows that $\|G_1(x)\| < 1 + \varepsilon$ and $\|F(x)\| \le 1$ for all $x \in D^n$. Thus F maps D^n into D^n and is C^∞. For $x \in D^n$,

$$\begin{aligned}\|G(x) - F(x)\| &= \|G(x) - (1 + \varepsilon)^{-1} G_1(x)\| \\ &= (1 + \varepsilon)^{-1} \|\varepsilon G(x) + G(x) - G_1(x)\| \\ &\le \varepsilon\|G(x)\| + \|G(x) - G_1(x)\| = 2\varepsilon.\end{aligned}$$

From these inequalities we obtain a contradiction to the assumption that every C^∞ map $F: D^n \to D^n$ leaves some point fixed. Namely, for every $x \in D^n$ we have

$$\begin{aligned}\|F(x) - x\| &= \|(G(x) - x) - (G(x) - F(x))\| \\ &\ge \|G(x) - x\| - \|G(x) - F(x)\| \\ &\ge 3\varepsilon - 2\varepsilon = \varepsilon.\end{aligned}$$

This contradiction shows that if every C^∞ map of D^n to D^n has a fixed point, then so must every continuous one. The proof of the theorem is then completed by the following lemma. ∎

(8.3) Lemma *If $F: D^n \to D^n$ is a C^∞ map, then F has a fixed point.*

Proof This is again a proof by contradiction. We suppose there exists an $F: D^n \to D^n$ which is C^∞ and has no fixed point. We shall use F to construct a C^∞ map from $\tilde{F}: D^n \to \partial D^n$ which leaves ∂D^n pointwise fixed. Namely, given $x \in D^n$, let $\tilde{F}(x)$ be the boundary point obtained by extending the segment $\overline{F(x)x}$ past x to the boundary of D^n (Fig. VI.13). In particular, if

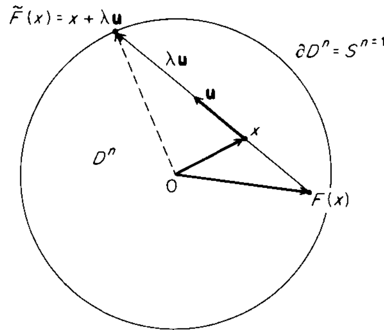


Figure VI.13

$x \in \partial D^n$, then $\tilde{F}(x) = x$ and, in any case, $\tilde{F}(D^n) \subset \partial D^n$. To see that $\tilde{F}$ is C^∞ we express $\tilde{F}$ explicitly using vector notation in $\boldsymbol{R}^n$. Namely, $\tilde{F}(x) = x + \lambda\mathbf{u}$, where x denotes the vector from $(0, \ldots, 0)$ to $x = (x^1, \ldots, x^n)$, $\mathbf{u}$ is the unit vector directed from $F(x)$ to x and lying on this segment, more precisely,

$$\mathbf{u} = \frac{x - F(x)}{\|x - F(x)\|},$$

and where $\lambda = -(x, \mathbf{u}) + [1 - (x, x) + (x, \mathbf{u})^2]^{1/2}$ denotes the length of the vector on $\mathbf{u}$ with initial point x and terminal point $\tilde{F}(x)$ on ∂D^n. Since F is C^∞, it is easy to check that $\tilde{F}$ is C^∞. Figure VI.13 represents the intersection of D^n with the 2-plane determined by three points: the origin 0, x, and $F(x)$. The scalar λ is the unique nonnegative number such that $\|x + \lambda\mathbf{u}\| = 1$. Since F is C^∞, $\mathbf{u}$ is C^∞, so wherever the expression under the radical is positive, then $\tilde{F}(x)$ is also C^∞. However, $1 - (x, x) \geq 0$ with equality only if $x \in S^{n-1}$; and $(x, \mathbf{u})^2 \geq 0$ with equality only when $\mathbf{u}$ is orthogonal to x, that is, when $x - F(x)$ is orthogonal to x. However, $(x, \mathbf{u}) = 0$ cannot occur when $(x, x) = 1$, that is, on a point of S^{n-1}, since in this case $F(x)$ would be exterior to D^n. Thus $1 - (x, x) + (x, \mathbf{u})^2 > 0$ on D^n and $\tilde{F}$ is C^∞. The existence of $\tilde{F}$ contradicts Lemma 8.1, so F has a fixed point, which completes the proof of Brouwer's fixed point theorem. ∎

As another application of these ideas we prove the following theorem concerning the antipodal map $A(x) = -x$ on the unit sphere S^{n-1} of $\boldsymbol{R}^n$.

(8.4) Theorem *If n is odd, then there is no C^∞ homotopy between the antipodal map $A: S^{n-1} \to S^{n-1}$ and the identity map of S^{n-1}.*

Proof The sphere is an orientable manifold, in fact we may define the oriented orthonormal frames of $T_x(S^{n-1})$ at each $x \in S^{n-1}$ in the following fashion. Each $x \in S^{n-1}$ determines a unit vector $\mathbf{x} = \overline{0x}$, and the elements of $T_x(S^{n-1})$ correspond to the vectors in the orthogonal complement of $\mathbf{x}$. If

$\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ is an orthonormal frame of $T_x(S^{n-1})$ in the induced metric of $\boldsymbol{R}^n$, then $\mathbf{x}, \mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ is an orthonormal frame of $\boldsymbol{R}^n$—we use the natural parallelism to identify vectors at distinct points of $\boldsymbol{R}^n$. Two frames, $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_{n-1}$, at x will be said to have the *same* orientation if the corresponding frames $\mathbf{x}, \mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ and $\mathbf{x}, \mathbf{e}'_1, \ldots, \mathbf{e}'_{n-1}$ do. Then from the canonical orientation of $\boldsymbol{R}^n$ we obtain an orientation of S^{n-1} by choosing as oriented that class of frames for which $\mathbf{x}, \mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ is an oriented frame of $\boldsymbol{R}^n$. This is exactly the orientation $\tilde{\partial} D^n$ introduced in connection with Stokes's theorem. Let Ω be the unique $(n-1)$-form on S^{n-1} which takes the value $+1$ on all oriented orthonormal frames $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$. Since $A: S^{n-1} \to S^{n-1}$ is the restriction to S^{n-1} of a linear, in fact an orthogonal, map of $\boldsymbol{R}^n$ its Jacobian is constant and just the map A itself. Thus under A the frame $\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ at x goes to the frame $-\mathbf{e}_1, \ldots, -\mathbf{e}_{n-1}$ at $-x$. It is clear that this will be oriented according to our orientation of S^{n-1} if and only if n is even so that $\mathbf{x}, \mathbf{e}_1, \ldots, \mathbf{e}_{n-1}$ and $-\mathbf{x}, -\mathbf{e}_1, \ldots, -\mathbf{e}_{n-1}$ are coherently oriented frames of $\boldsymbol{R}^n$. Therefore $A^*\Omega = (-1)^n\Omega$ and when n is odd, $\Omega = -A^*\Omega$.

If there were a C^∞ homotopy connecting A and the identity, then $\Omega - A^*\Omega$ must be exact by Theorem 7.13. Since the integral over S^{n-1} of an exact form is zero by Stokes's theorem, this means that when n is odd,

$$2\int_{S^{n-1}} \Omega = \int_{S^{n-1}} (\Omega - A^*\Omega) = 0.$$

But this is impossible since the integral over S^{n-1} of the volume element is positive. [For an alternative proof see Exercise 3.] ∎

We shall deduce two consequences. First, recall that although we have discussed orientability of manifolds at some length and have shown that some manifolds are orientable, for example, $S^{n-1} = \partial D^n$, we have never presented an example of a manifold which cannot be oriented. We shall remedy this omission now.

(8.5) Corollary *Real projective space $P^n(\boldsymbol{R})$ is not orientable when n is even.*

Proof Suppose that it is; we know that S^n is a (two-sheeted) covering manifold of $P^n(\boldsymbol{R})$ which can thus be obtained from S^n as the orbit space of the group of two elements $\boldsymbol{Z}_2$ acting on S^n. This action is obtained by letting the generator of $\boldsymbol{Z}_2$ correspond to the antipodal map A (Example III.8.2). If Ω is a nowhere vanishing n-form on $P^n(\boldsymbol{R})$ and $F: S^n \to P^n(\boldsymbol{R})$ is the covering map, then $F^*\Omega = \Omega^*$ is a nowhere vanishing n-form on S^n. Moreover since $F \circ A = F$ we see that $A^*\Omega^* = \Omega^*$, which is not possible since A reverses orientation if $n + 1$ is odd. Thus $P^n(\boldsymbol{R})$ is not orientable when n is even. ∎

(8.6) Theorem *If n is even, then there does not exist a C^∞-vector field X on S^n which is not zero at some point.*

Proof We suppose that such a vector field exists and show that this implies that the antipodal map A and the identity map I on S^n are C^∞ homotopic. Let X be a C^∞-vector field on S^n such that X is never zero. Then $X/\|X\|$ is a C^∞-vector field of unit vectors (we use the induced metric of $\boldsymbol{R}^{n+1}$) so we may suppose to begin with that $\|X\| \equiv 1$ on S^n. If x is a point of S^n, let X_x be the corresponding vector of the field. Treating $\boldsymbol{R}^{n+1}$ as a vector space and thinking of x as a radius vector, we have $(x, X_x) = 0$ for every x, and we define the homotopy $H: S^n \times I \to S^n$ by

$$H(x, t) = (\cos \pi t)x + (\sin \pi t)X_x .$$

Then $H(x, t)$ is C^∞ and, since $\|H(x, t)\| \equiv 1$, it defines a map of $S^n \to S^n$ for each t. Thus $H(x, 0) \equiv x$ and $H(x, 1) \equiv -x$ as claimed. However, the existence of such a homotopy when n is even contradicts the previous proposition; therefore in this case no such vector field exists. ∎

(8.7) Remark We have noted previously that when n is odd, then the vector field X_x assigning to $x = (x^1, x^2, \ldots, x^n, x^{n+1}) \in S^n$ the unit vector

$$X_x = x^2 \frac{\partial}{\partial x^1} - x^1 \frac{\partial}{\partial x^2} + \cdots + x^{n+1} \frac{\partial}{\partial x^n} - x^n \frac{\partial}{\partial x^{n+1}}$$

orthogonal to x defines a nowhere vanishing field of tangent vectors to S^n. It follows that in this case A is homotopic to the identity.

The de Rham Groups of Lie Groups

We shall briefly touch on a special case of considerable interest in the theory of de Rham groups; in fact it is the case which may have led to their discovery. We suppose that G is a compact connected Lie group, for example $SO(n)$, and that $\theta: G \times M \to M$ is an action of G on a compact manifold M. As usual θ_g denotes the diffeomorphism of M defined by $\theta_g(p) = \theta(g, p)$. A covariant tensor φ on M, in particular an exterior differential form, is said to be *invariant* if $\theta_g^*\varphi = \varphi$ for each $g \in G$. Since $d(\theta_g^*\varphi) = \theta_g^*(d\varphi)$ for every form φ (Theorem V.5.2), we see that if φ is invariant, $d\varphi$ is also. Let $\tilde{\bigwedge}^k(M)$ denote the subspace of $\bigwedge^k(M)$ which consists of all invariant k-forms. Then $d(\tilde{\bigwedge}^k(M)) \subset \tilde{\bigwedge}^{k+1}(M)$ as we have just seen; and we may define $\tilde{Z}^k(M) = \{\varphi \in \tilde{\bigwedge}^k(M) \mid d\varphi = 0\}$ and $\tilde{B}^k(M) = d(\tilde{\bigwedge}^k(M)) \subset \tilde{Z}^k(M)$, the closed invariant forms and "invariantly exact" forms of degree k. We then make the following definition.

(8.8) Definition The *invariant de Rham groups* of M, denoted by $\tilde{H}^k(M)$, are defined by $\tilde{H}^k(M) = \tilde{Z}^k(M)/\tilde{B}^k(M)$.

We note that the natural inclusion i of $\tilde{\bigwedge}^k(M)$ in $\bigwedge^k(M)$ takes $\tilde{Z}^k(M)$ into $Z^k(M)$ and $\tilde{B}^k(M)$ into $B^k(M)$ and hence induces a homomorphism

$i_*: \tilde{H}^k(M) \to H^k(M)$. In order to study this homomorphism we define an $\boldsymbol{R}$-linear operator $\mathscr{P}: \bigwedge^k(M) \to \tilde{\bigwedge}^k(M)$. If $\varphi \in \bigwedge^k(M)$, then let Ω denote the bi-invariant volume element for which $\text{vol}(G) = 1$ and define $\mathscr{P}\varphi$ by

$$\mathscr{P}\varphi(X_1, \ldots, X_k) = \int_G \theta_g^*\varphi(X_1, \ldots, X_k)\Omega.$$

This operator has the following properties:

(8.9) Lemma *$\mathscr{P}$ takes a k-form to an invariant k-form, that is, $\mathscr{P}(\bigwedge^k(M)) \subset \tilde{\bigwedge}^k(M)$. Moreover,*

(i) *if $\varphi \in \tilde{\bigwedge}^k(M)$, then $\mathscr{P}\varphi = \varphi$;*
(ii) *$d\mathscr{P} = \mathscr{P}d$.*

Proof It is easy to check that $\mathscr{P}\varphi \in \tilde{\bigwedge}^k(M)$ and in fact is G-invariant:

$$\theta_a^*\mathscr{P}\varphi(X_1, \ldots, X_k) = \mathscr{P}\varphi(\theta_{a*}X_1, \ldots, \theta_{a*}X_k) = \int_G \theta_g^*\varphi(\theta_{a*}X_1, \ldots, \theta_{a*}X_k)\Omega$$
$$= \int_G \theta_a^*[\theta_g^*\varphi(X_1, \ldots, X_k)]\Omega = \int_G \theta_{ga}^*\varphi(X_1, \ldots, X_k)\Omega$$
$$= \int_G \theta_g^*\varphi(X_1, \ldots, X_k)\Omega.$$

The fact that $\mathscr{P}\varphi$ is C^∞ and (ii) are consequences of the Leibniz rule for differentiating under the integral sign. If φ is G-invariant, then $\theta_g^*\varphi = \varphi$ for all $g \in G$, or more precisely at each $p \in M$,

$$\theta_g^*\varphi_{\theta(g,p)}(X_{1p}, \ldots, X_{kp}) = \varphi_p(X_{1p}, \ldots, X_{kp}).$$

From this it follows that

$$\mathscr{P}\varphi(X_1, \ldots, X_k) = \int_G \theta_g^*\varphi(X_1, \ldots, X_k)\Omega = \varphi(X_1, \ldots, X_k)\int_G \Omega.$$

Since $\int_G \Omega = 1$, $\mathscr{P}\varphi = \varphi$ and property (i) are established. ∎

The lemma leads to the following result for G, M and $\tilde{H}^k(M)$ as described above:

(8.10) Theorem *The homomorphism $i_*: \tilde{H}^k(M) \to H^k(M)$ is an isomorphism into for each $k = 0, 1, \ldots, \dim M$.*

Proof Suppose that $[\tilde{\varphi}]$ is an element of $\tilde{H}^k(M)$ and that $\tilde{\varphi}$ is a closed invariant form on M belonging to the class $[\tilde{\varphi}]$. In order to show that i_* is one-to-one we need only see that if $\tilde{\varphi} = d\sigma$, $\sigma \in \bigwedge^{k-1}(M)$, then $\tilde{\varphi}$ is the image under d of an element of $\tilde{\bigwedge}^{k-1}(M)$, that is, that if $\tilde{\varphi}$ is exact, then it is "invariantly exact." This follows from Lemma 8.9 since $\tilde{\varphi} = \mathscr{P}\tilde{\varphi} = \mathscr{P}\, d\sigma = d(\mathscr{P}\sigma)$ and $\mathscr{P}\sigma \in \tilde{\bigwedge}^{k-1}(M)$. ∎

(8.11) Remark It is also true, but somewhat harder to prove directly, that i_* is *onto*, that is, $\tilde{H}^k(M)$ is isomorphic to $H^k(M)$ for all k. For details the reader is referred to Chevalley and Eilenberg [1] or Greub *et al.* [1].

This theorem, coupled with the remark, enables one to reduce to algebraic form many questions about the cohomology of homogeneous spaces. In particular, computation of the de Rham groups of such a space can be converted to a problem concerning representations of the isotropy subgroup of a point. Consideration of these interesting questions would take us too far afield, however we touch very briefly on one special case.

(8.12) Lemma *Let Φ_e be a covariant tensor of order r on $T_e(G)$, where G is a connected Lie group. If* $\operatorname{Ad} g^*\Phi_e = \Phi_e$, *that is, if Φ_e determines a bi-invariant tensor on G, then for any $X_1, \ldots, X_r, Z \in \mathfrak{g}$, we have*

$$\sum_{i=1}^{r} \Phi(X_1, \ldots, [Z, X_i], \ldots, X_r) = 0.$$

Proof Let Φ denote the bi-invariant covariant tensor on G determined by Φ_e. Given $Z \in \mathfrak{g}$, a left-invariant vector field on G, we have seen (Sections IV.5 and IV.6) that Z is complete and that the one-parameter group action $\theta\colon R \times G \to G$ which it determines is given by right translations by the elements of a uniquely determined one-parameter subgroup $g(t) = \exp tZ$ by the formula $\theta_t = R_{g(t)}$. We have previously (Theorem IV.7.8) established the following formula for C^∞-vector fields on a manifold (in this case on G):

$$[Z, X]_p = \lim_{t \to 0} \frac{1}{t} [\theta_{-t*} X_{\theta_t(p)} - X_p].$$

If we suppose that $p = e$ and that X is a left-invariant vector field, then $[Z, X]$ is just the product in the Lie algebra $\mathfrak{g}$. If we identify $\mathfrak{g}$ with $T_e(G)$, we may write

$$[Z, X] = \lim_{t \to 0} \frac{1}{t} [R_{g(-t)*} X_{g(t)} - X_e].$$

Since Φ is bi-invariant, $R^*_{g(-t)}\Phi - \Phi = 0$; thus for any $X_1, \ldots, X_r \in \mathfrak{g}$,

$$\Phi(R_{g(-t)*}X_1, \ldots, R_{g(-t)*}X_r) - \Phi(X_1, \ldots, X_r) = 0.$$

Adding and subtracting $\Phi(X_1, \ldots, X_{i-1}, R_{g(-t)*}X_i, \ldots, R_{g(-t)*}X_r)$, $i = 1, \ldots, r$, then multiplying by $1/t$ and letting $t \to 0$, we obtain the formula.

(8.13) Corollary *Every bi-invariant exterior form on a Lie group G is closed.*

Proof Let ω be an exterior differential r-form. If ω is left-invariant and $X_1, X_2, \ldots, X_{r+1}$ are left-invariant, then, using a caret to denote missing terms,

$$d\omega(X_1, \ldots, X_{r+1}) = \sum_{i=1}^{r+1} (-1)^{i-1} X_i \omega(X_1, \ldots, \hat{X}_i, \ldots, X_{r+1})$$
$$+ \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{r+1}).$$

We previously established this formula for $r = 1$ (Lemma V.8.4). The method of proof in the general case is the same (compare Exercise V.8.3). The first term on the right vanishes since ω and X_j, $j = 1, \ldots, r+1$, are left-invariant, the second as a consequence of the lemma. ∎

Now we suppose that G acts on itself by both left and right translations. More formally, let $G = M$ and let $K = G \times G$, the direct product of Lie groups, and define $\theta\colon K \times M \to M$ for each $x \in M = G$ and $k = (g_1, g_2) \in K$ by

$$\theta(k, x) = g_1 x g_2^{-1} \quad (= R_{g_2^{-1}} \circ L_{g_1}(x)).$$

Then the K-invariant forms $\tilde{\varphi}$ on G are exactly the bi-invariant forms.

(8.14) Corollary *Each bi-invariant r-form on a compact, connected, Lie group G determines a nonzero element of $H^r(G)$.*

Proof By Corollary 8.13 each $\tilde{\varphi} \in \tilde{H}^r(G)$, that is, each bi-invariant r-form is closed. We know that if it is exact, then it must be of the form $d\tilde{\sigma}$ with $\tilde{\sigma}$ bi-invariant. But then it is zero, by the corollary again, since $d\tilde{\sigma} = 0$. ∎

(8.15) Example Consider any compact, connected, *non-Abelian* Lie group G, for example, $SO(n)$, the orthogonal matrix group (with elements of determinant $+1$) for $n \geq 3$. We claim that $H^3(G) \neq \{0\}$. We consider that the exterior three-form $\varphi(X, Y, Z) = ([X, Y], Z)$ on G; (X, Y) denotes the bi-invariant inner product. Since $X, Y \in \mathfrak{g}$ implies that $[X, Y]$ is left-invariant and since $\mathrm{Ad}(g)$ is an automorphism of $\mathfrak{g}$, it follows readily that φ is bi-invariant. The alternating property of φ follows from $[X, Y] = -[Y, X]$ together with the symmetry of (X, Y). By Corollary 8.14, φ is closed and if it is not zero, it determines an element of $H^3(G)$. Suppose that $\varphi = 0$. Then for all $X, Y, Z \in \mathfrak{g}$, we have $\varphi(X, Y, Z) = ([X, Y], Z) = 0$. In particular, $([X, Y], [X, Y]) = 0$ so that $[X, Y] = 0$ for all $X, Y \in \mathfrak{g}$. This means, according to Section IV.7, that the one-parameter groups of G commute. It follows that there is a neighborhood U of e which consists of commuting elements. By the connectedness of G it follows that the elements of U generate G, which is therefore commutative, contrary to assumption. This means that φ determines a nonvanishing element $[\varphi]$ of $H^3(G)$.

Exercises

1. Let A be a closed subset of the metric space $\boldsymbol{R}^n$ and $f: A \to \boldsymbol{R}$ a continuous bounded function. Show that there is a continuous extension of f to all of $\boldsymbol{R}^n$ (Tietze–Urysohn extension theorem).
2. Given $\delta > 0$, let $g_\delta(x)$ be a nonnegative C^∞ function on $\boldsymbol{R}^n$ such that $\operatorname{supp} g_\delta \subset B_\delta^n(0)$ and $\int_{\boldsymbol{R}^n} g_\delta(x)\, dx^1 \cdots dx^n = 1$. Let f be a continuous function defined on an open subset U and let K be compact with $K \subset U$. Given $\varepsilon > 0$, choose $\delta > 0$ so that: (i) δ is less than the distance from K to $\boldsymbol{R}^n - U$ and (ii) for $\|y\| < \delta$ and $x \in K$, $|f(y + x) - f(x)| < \varepsilon$. Show that
$$g(x) = \int_{\boldsymbol{R}^n} f(y + x) g_\delta(y)\, dy^1 \cdots dy^n$$
is a C^∞ function and that $|g(x) - f(x)| < \varepsilon$ on K.
3. Prove Theorem 8.4 by using on S^{n-1} the form Ω obtained by restricting $\sum_{j=1}^n (-1)^{j-1} x^j\, dx^1 \wedge \cdots \wedge dx^{j-1} \wedge dx^{j+1} \wedge \cdots \wedge dx^n$ to S^{n-1}.
4. Show that if n is odd, then $P^n(\boldsymbol{R})$ is orientable.
5. Show that there exists no *continuous* vector field on S^n, n even, which is nowhere zero.
6. Using Remark 8.11, prove that if G is a connected compact Lie group of dimension n, then $H^n(G) \cong \boldsymbol{R}$.
7. Using Remark 8.11 compute $H^k(T^n)$, $k = 0, 1, \ldots, n$. Give a formula for $\dim H^k(T^n)$ (as a vector space over $\boldsymbol{R}$).
8. Define the *center* $\mathfrak{c}$ of the Lie algebra $\mathfrak{g}$ of the compact connected Lie group G by $\mathfrak{c} = \{Z \in \mathfrak{g} \mid [Z, X] = 0\ \forall X \in \mathfrak{g}\}$. Prove that it is a subalgebra and that if $Z \in \mathfrak{c}$, then $\exp tZ$ is in the center of G. Show that if $\mathfrak{c} = \{0\}$, then $\tilde{H}^1(G) = \{0\}$.

9 Covering Spaces and the Fundamental Group

The ideas involved in paths, loops, their homotopies, and the fundamental group of a manifold M, which were discussed in Section 6 have an intimate connection with the covering spaces of M and with properly discontinuous groups, which were considered much earlier (Sections III.8 and III.9). Clarifying this relationship will enable us to complete the discussion of Chapter III in several important respects.

Suppose in what follows that M is a manifold, $\tilde{M}$ a covering manifold, and $F: \tilde{M} \to M$ the (C^∞) covering mapping. If X is a topological space and $G: X \to M$ a continuous mapping, then a continuous mapping $\tilde{G}: X \to \tilde{M}$ is said to *cover* G if $F \circ \tilde{G} = G$; we also say $\tilde{G}$ is a *lift* of G. For example, if $f: I \to M$ is a path or loop, then $\tilde{f}: I \to \tilde{M}$ is a path which covers it if $F \circ \tilde{f}(t) = f(t)$ for $0 \le t \le 1$. If a covering $\tilde{f}$ of a given path f exists at all, then it is uniquely determined by its value on a single point, say by $f(0)$. More generally, with the notation above we have the following results.

(9.1) Lemma *If $F: \tilde{M} \to M$ is a covering and X is a connected space, then two (continuous) mappings $\tilde{G}_1, \tilde{G}_2: X \to \tilde{M}$ covering a continuous mapping $G: X \to M$ agree if they have the same value at a single point $x_0 \in X$.*

Proof Let $A = \{x \in X: \tilde{G}_1(x) = \tilde{G}_2(x)\}$. Then A is closed by continuity of $\tilde{G}_1$ and $\tilde{G}_2$. A is also open, for if $x \in A$ and if U is a neighborhood of $\tilde{G}_1(x) = \tilde{G}_2(x)$ such that $F \mid U$ is a diffeomorphism of U to M, then it follows that $\tilde{G}_1$ and $\tilde{G}_2$ must agree on the open set $V = \tilde{G}_1^{-1}(U) \cap \tilde{G}_2^{-1}(U)$. In fact if $y \in V$, then $F \circ \tilde{G}_1(y) = F \circ \tilde{G}_2(y)$ by hypothesis; but since $\tilde{G}_1(y)$ and $\tilde{G}_2(y)$ are in U, on which F is one-to-one, they must be equal. Finally since A is not empty and X is connected, $A = X$. ∎

(9.2) Theorem *Let $f: I \to M$ be a path in M with initial point $b = f(0)$. If $F: \tilde{M} \to M$ is a covering and $\tilde{b} \in F^{-1}(b)$, then there is a unique path $\tilde{f}$ in $\tilde{M}$ with initial point $\tilde{f}(0) = \tilde{b}$ which covers $f(t)$.*

Proof Uniqueness is a consequence of the previous proposition. To prove existence we suppose $0 = t_0 < t_1 < \cdots < t_n = 1$ is any partition of I such that for each $i, f([t_i, t_{i+1}])$ lies in an admissible neighborhood V_i with respect to the covering. The existence of such a partition follows from the compactness of I and the continuity of f. We let $f(0) = b$ and let $\tilde{b} \in \tilde{M}$ denote a point over b, that is, $F(\tilde{b}) = b$. If U_1 is the unique connected component of $F^{-1}(V_1)$ containing $\tilde{b}$, then we define $\tilde{f}(t)$, $0 \le t \le t_1$, by $\tilde{f}(t) = (F \mid U_1)^{-1}(f(t))$. Then $\tilde{f}(t_1) \in U_1 \cap U_2$, where U_2 is the unique component of $F^{-1}(V_2)$ containing $\tilde{f}(t_1)$. This allows us to define $f(t) = (F \mid U_2)^{-1}(f(t))$ for $t_1 \le t \le t_2$ and thus determine $\tilde{f}$ on $[t_0, t_2]$. Clearly we can continue in this fashion to define $\tilde{f}$ on all of I. (See Fig. VI.14.) ∎

It is important to realize that not only a path but even a homotopy of paths can be lifted to a covering $\tilde{M}$ of M by this procedure.

Suppose that $f, g: I \to M$ are paths and $H: I \times I \to M$ is a homotopy of f and g, i.e., $H(0, s) \equiv f(s)$, $H(1, s) \equiv g(s)$, and $F: \tilde{M} \to M$ is a covering of M as above.

(9.3) Theorem *Given $\tilde{b} \in \tilde{M}$ such that $F(\tilde{b}) = b = f(0)$ $(= H(0, 0)$, then there is a uniquely determined homotopy $\tilde{H}(t, s)$ with $\tilde{H}(0, 0) = \tilde{b}$ which covers $H(t, s)$, i.e., such that $F \circ \tilde{H} = H$. If H leaves fixed the endpoints of f and g, then $\tilde{H}$ leaves fixed the endpoints of the paths $\tilde{H}_t$ covering $H_t(s)$ at each stage of the homotopy.*

Proof First suppose that $H(I \times I) \subset V$, an admissible neighborhood on M relative to the covering. Let $U \subset \tilde{M}$ be the unique neighborhood in $F^{-1}(V)$ which contains $\tilde{b}$. Then $F|U: U \to V$ is a diffeomorphism and

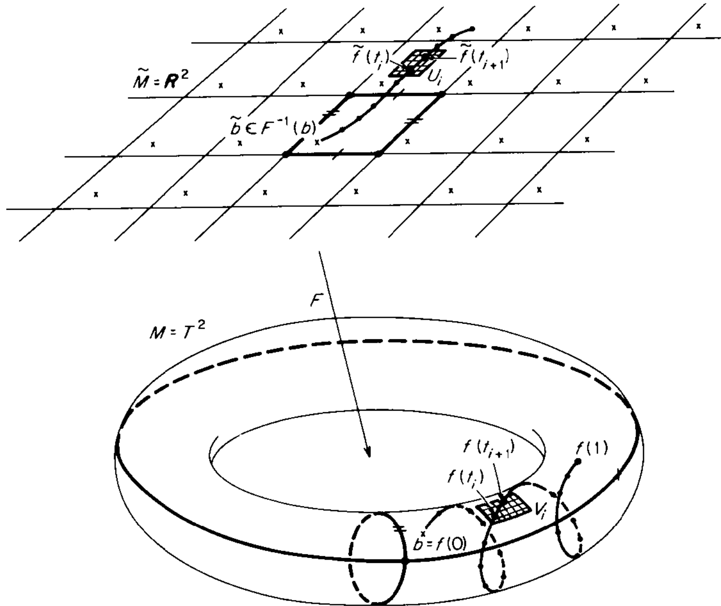


Figure VI.14

Lifting a path from T^2 to its covering R^2.

$\tilde{H} = (F|U)^{-1} \circ H$ is the unique lift of H such that $\tilde{H}(0, 0) = \tilde{b}$. In the general case we choose $\varepsilon = 1/k$ sufficiently small so that by partitioning $I \times I$ into k^2 squares of side ε, i.e., partitioning by lines $t = i/k$, $s = j/k$, $i, j = 0, 1, \ldots, k$, we ensure that each square $Q_{ij} = \{(t, s) | i/k \leq t \leq (i + 1)/k, j/k \leq s \leq (j + 1)/k\}$, $0 \leq i, j \leq k - 1$, lies in an admissible neighborhood V_{ij}. This is possible since $H(I \times I)$ is compact. We then proceed as follows: Define $\tilde{H}_{00}: Q_{00} \to \tilde{M}$ so that $\tilde{H}_{00}(0, 0) = \tilde{b}$ and $F \circ \tilde{H}_{00} = H|Q_{00}$, using the method above. This lift is unique and defines $\tilde{H}_{00}(0, 1/k)$ so we may define $\tilde{H}_{01}: Q_{01} \to \tilde{M}$ so that $\tilde{H}_{01}(0, 1/k) = \tilde{H}_{00}(0, 1/k)$ and $F \circ \tilde{H}_{01} = H|Q_{01}$. By uniqueness, $\tilde{H}_{00} = \tilde{H}_{01}$ on their common boundary $Q_{00} \cap Q_{01}$. Continuing in this way we define $\tilde{H}_{00}, \tilde{H}_{01}, \ldots, \tilde{H}_{0k-1}$.

This defines a unique continuous lift $\tilde{H}_0$ of H restricted to the strip $\{(t, s) | 0 \leq t \leq 1/k,\ 0 \leq s \leq 1\}$, with $\tilde{H}_0(0, 0) = \tilde{b}$. We now continue this process on the next row of squares: $Q_{10}, \ldots, Q_{1k-1}$, etc. (see Fig. VI.15). After a finite number of steps, a continuous mapping $\tilde{H}: I \times I \to M$ is defined such that $\tilde{H}(0, 0) = \tilde{b}$ and $F \circ \tilde{H} = H$.

If the endpoints remain fixed at all stages of the homotopy H, then we must have $H(t, 0) = b$ and $H(t, 1) = c$, common initial and terminal points of f, g. Since $F \circ \tilde{H}(t, 0) = b$, it follows that the continuous path $t \to \tilde{H}(t, 0)$ lies in the discrete set $F^{-1}(b)$. Hence it must be a single point—in this case $\tilde{b}$. Similarly, $\tilde{H}(t, 1)$ is a single point lying over c.

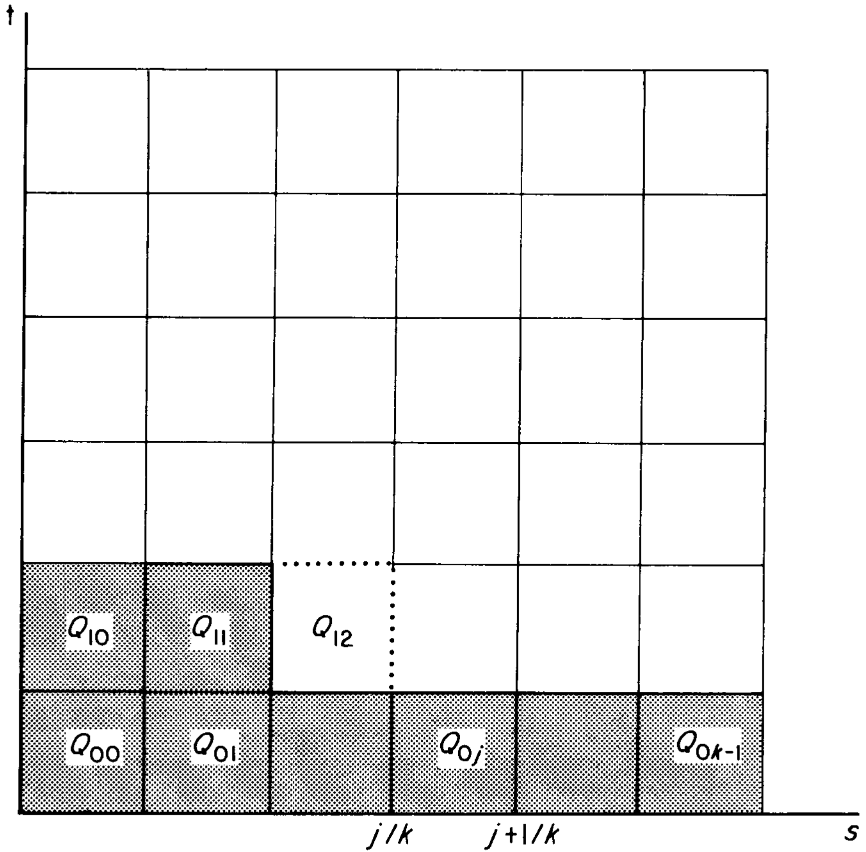


Figure VI.15
The next step extends the lift $\tilde{H}$ of H from shaded squares to Q_{12}.

(9.4) Corollary *If $\tilde{b} \in \tilde{M}$ lies over $b \in M$, then $F_*: \pi_1(\tilde{M}, \tilde{b}) \to \pi_1(M, b)$ is an injective isomorphism.*

Proof We know F_* is a homomorphism and, using the previous theorem with $\tilde{f}$, $\tilde{g}$ loops at $\tilde{b}$, we see that $F \circ \tilde{f} \sim F \circ \tilde{g}$ implies $\tilde{f} \sim \tilde{g}$. This is equivalent to F_* being injective. ∎

We conclude this section by proving two theorems which give a much more precise picture of the relation between coverings of a manifold M and the fundamental group. They will also enable us to complete our discussion of the relation between covering spaces and orbit spaces of a properly discontinuous action of a group Γ on a manifold, which was considered in Section III.9. If $\tilde{M}_1$ and $\tilde{M}_2$ are coverings of a manifold M with covering

maps F_1 and F_2, then a homeomorphism $G: \tilde{M}_1 \to \tilde{M}_2$ such that $F_1 = F_2 \circ G$ and $F_2 = F_1 \circ G^{-1}$ is called an *isomorphism* of the coverings. In particular, an automorphism, that is, isomorphism, $G: \tilde{M} \to \tilde{M}$ is exactly a covering transformation, as given in Definition III.9.2. Using admissible neighborhoods, it is apparent that the differentiability of F_1 and F_2 implies that of G and G^{-1}. We now show that in a sense isomorphism classes of coverings of M are in one-to-one correspondence with subgroups of the fundamental group.

(9.5) Theorem *Let $F_1: \tilde{M}_1 \to M$ and $F_2: \tilde{M}_2 \to M$ be coverings of the same manifold M. Suppose that for $b \in M$, $\tilde{b}_1 \in \tilde{M}_1$ and $\tilde{b}_2 \in \tilde{M}_2$ such that $F_1(\tilde{b}_1) = b = F_2(\tilde{b}_2)$ we have $F_{1*}\pi_1(\tilde{M}_1, \tilde{b}_1) = F_{2*}\pi_1(\tilde{M}_2, \tilde{b}_2)$ [as subgroups of $\pi_1(M, b)$]. Then there is exactly one isomorphism $G: \tilde{M}_1 \to \tilde{M}_2$ taking $\tilde{b}_1$ to $\tilde{b}_2$. Conversely, if such a G exists, then $F_{1*}(\pi_1(\tilde{M}_1, \tilde{b}_1) = F_{2*}(\pi_2(\tilde{M}_2, \tilde{b}_2)$.*

Proof Given $\tilde{p} \in \tilde{M}_1$ we define $G(\tilde{p})$ as follows: Let $\tilde{f}_1$ be a path such that $\tilde{f}_1(0) = \tilde{b}$ and $\tilde{f}_1(1) = \tilde{p}$. Then the path $f = F_1 \circ \tilde{f}_1$ on M has a unique lifting to a path $\tilde{f}_2$ on $\tilde{M}_2$ covering f and with initial point $\tilde{f}_2(0) = \tilde{b}_2$. We define $G(\tilde{p}) = \tilde{f}_2(1)$. Of course we must show that the definition is independent of the path $\tilde{f}_1$ chosen, and that G is continuous. On the other hand, once these facts are proved, then we see that $F_1 = F_2 \circ G$ and that $G(\tilde{b}_1) = \tilde{b}_2$ are immediate consequences of the definition. Uniqueness of G will follow as well since any G with the properties required in the theorem must take $\tilde{f}_1$ to a path $\tilde{f}_2$ on $\tilde{M}_2$ which covers $f = F_1 \circ \tilde{f}_1$ and runs from $\tilde{b}_2$ to $G(\tilde{p})$. This applies to loops as well, which proves the converse.

Now suppose that $\tilde{f}_1$ and $\tilde{g}_1$ are distinct paths on $\tilde{M}_1$ from $\tilde{b}_1$ to $\tilde{p}$. Let $f = F_1 \circ \tilde{f}_1$ and $g = F_1 \circ \tilde{g}_1$ and consider the loop $f * g^{-1}$ with $g^{-1}(s) = g(1-s)$, $0 \le s \le 1$. This loop determines an element $[f * g^{-1}]$ of $F_{1*}\pi_1(\tilde{M}_1, \tilde{b}_1)$ and hence also the (same) element of $F_{2*}\pi_1(\tilde{M}_2, \tilde{b}_2)$. In view of the preceding corollary, if we lift this to a path from $\tilde{b}_2$, its terminal point will necessarily be $\tilde{b}_2$; and so the lifted paths $\tilde{f}_2$ and $\tilde{g}_2$ on $\tilde{M}_2$ beginning at $\tilde{b}_2$ both end at the same point, that is, $\tilde{f}_2(1) = \tilde{g}_2(1)$. It follows that by using either $\tilde{f}_1$ or $\tilde{g}_1$ we obtain the same value for $G(\tilde{p})$. We also see from this line of argument that there is a one-to-one correspondence between *points* of $\tilde{M}_i$, $i = 1, 2$, and *equivalence classes of paths* f on M issuing from b relative to the subgroup $A = F_*(\pi_1(\tilde{M}_i, \tilde{b}_i) \subset \pi_1(M, b)$, where equivalence ($\approx$) relative to A is defined for paths f, g with $f(0) = b = g(0)$ by $f \approx g$ if $f(1) = g(1)$ and $[f * g^{-1}] \in A$. In fact, let $p \in M$, $\{f\}$ an equivalence class of paths from b to p; $\{f\}$ determines a point $\tilde{p}_{\{f\}}$ of M_1 which lies over p. Indeed, the class $\{f\}$ lifts to a class $\{\tilde{f}\}$, all curves of which issue from the point $\tilde{b}_1$; and we have just seen that they all have as terminal point $\tilde{p}$. Conversely, when $F: \tilde{M} \to M$ is a covering, fixing b, $\tilde{b} \in F^{-1}(b)$ defines a correspondence of points $\tilde{p} \in \tilde{M}$ and equivalence classes as follows: Let $\tilde{f}$ be a path from $\tilde{b}$ to $\tilde{p}$;

then $\tilde{p} \leftrightarrow \{F \circ \tilde{f}\}$. The classes of loops at b are then in one-to-one correspondence with the points over b. These remarks should clarify the intuitive situation and are useful later. The reader might find it helpful to think of the example $\tilde{M} = \mathbf{R}^2$, $M = T^2$.

To continue the proof, it is clear that G is onto and has an inverse G^{-1} which is described in a completely symmetrical way to G. Now let $\tilde{p}_2 = G(\tilde{p}_1) \in \tilde{M}_2$ and let V, ψ be an admissible coordinate neighborhood of $p = F_i(\tilde{p}_i)$ on M, $i = 1, 2$. We also suppose $\psi(V) = B_1^n(0) \subset \mathbf{R}^n$ and $\psi(p) = 0$. If f is a path from b to p on M which lifts to paths $\tilde{f}_i$ joining $\tilde{b}_i$ to $\tilde{p}_i$ on $\tilde{M}_i$, $i = 1, 2$, then we see that this path may be used in the definition of G as described above. For any point q in V we have a radial path (in the local coordinates), say g_q, from p to q and $f_q = f * g_q$ lifts to paths from $\tilde{b}_i$ to $\tilde{q}_i$ in the component $\tilde{U}_i$ of $F_i^{-1}(V)$ containing $\tilde{b}_i$, $i = 1, 2$; thus $G(\tilde{q}_1) = \tilde{q}_2$. This description being unique, and valid for every $q \in V$, we see that $G\colon \tilde{U}_1 \to \tilde{U}_2$ is one-to-one and onto and in fact may be described as follows: $G \mid \tilde{U}_1 = (F_2 \mid \tilde{U}_2)^{-1} \circ (F_1 \mid \tilde{U}_1)$. Thus $G \mid \tilde{U}_1$ is a diffeomorphism and since $\tilde{M}_1$ is covered by open sets of this type, G is differentiable, which completes the proof. ∎

(9.6) Corollary *If $F\colon \tilde{M} \to M$ is a covering and $\tilde{M}$ is simply connected, then the covering transformations are simply transitive on each set $F^{-1}(p)$. If we fix $\tilde{b} \in \tilde{M}$ and $b \in M$ with $F(\tilde{b}) = b$, then these choices determine a natural isomorphism $\Phi\colon \pi_1(M, b) \to \tilde{\Gamma}$ of the fundamental group of M onto the group of covering transformations.*

Proof Suppose that $q_1, q_2 \in F^{-1}(p)$. We apply Theorem 9.5 with $\tilde{M}_1 = \tilde{M}$, $\tilde{M}_2 = \tilde{M}$ to obtain a covering transformation $G\colon \tilde{M} \to \tilde{M}$ such that $G(q_1) = q_2$. To see that the hypotheses are satisfied it is only necessary to note that because $\tilde{M}$ is simply connected, $\pi_1(\tilde{M}, q_i) = \{1\}$, $i = 1, 2$; hence $F_*(\pi_1(\tilde{M}, q_1)) = \{1\} = F_*(\pi_1(\tilde{M}, q_2))$. We remark that by Theorem III.9.3, it now follows that the group $\tilde{\Gamma}$ of covering transformations must therefore be simply transitive on $F^{-1}(p)$ for each $p \in M$.

Having fixed $b \in M$ and $\tilde{b} \in \pi^{-1}(b)$, we may establish an isomorphism of $\pi_1(M, b)$ and $\tilde{\Gamma}$ as follows. Let $[g] \in \pi_1(M, b)$ and $\tilde{g}$ the lift of $g \in [g]$ to $\tilde{M}$ determined by $\tilde{g}(0) = \tilde{b}$. We have seen earlier that any two curves $\tilde{g}_1, \tilde{g}_2$ which are lifts of homotopic curves are homotopic. In particular two loops of $[g]$ with $\tilde{g}_1(0) = \tilde{b} = \tilde{g}_2(0)$, must have the same terminal point $\tilde{b}_1$ and must be homotopic (with endpoints fixed). Since g is a loop, $F(\tilde{b}) = b = F(\tilde{b}_1)$. We let $\Phi[g] \in \tilde{\Gamma}$ be the covering transformation taking $\tilde{b}$ to $\tilde{b}_1 = \tilde{g}(1)$. This defines $\Phi\colon \pi_1(M, b) \to \tilde{\Gamma}$. It is easily checked that Φ is a homomorphism using the arguments of the preceding theorem. If $\Phi[g] = 1$, then $\tilde{g}(0) = \tilde{b} = \tilde{g}(1)$ so that $\tilde{g}$ determines an element of $\pi_1(\tilde{M}, \tilde{b})$. Since this group contains only the identity, we have $\tilde{g} \sim e_{\tilde{b}}$ by a homotopy $\tilde{H}$. Then $H = F \circ \tilde{H}$ is a homotopy of g to e_b so $[g] = 1$ and hence Φ is one-to-one.

We also see that Φ is onto: if $G_1 \in \tilde{\Gamma}$, then let $\tilde{b}_1 = G_1(\tilde{b})$. There is a path $\tilde{g}$ from $\tilde{b}$ to $\tilde{b}_1$ and since $F(\tilde{b}) = F[G_1(\tilde{b})]$ by definition of covering transformation, $g = F \circ \tilde{g}$ is a loop at b. It determines $[g] \in \pi_1(M, b)$; and since the covering transformation $G = \Phi([f])$ agrees with G_1 on $\tilde{b}$, $G_1(\tilde{b}) = \tilde{b}_1 = G(\tilde{b})$, we must have $G = G_1$ by the results of Chapter III (or by Lemma 9.1). ∎

(9.7) Theorem *Let M be a connected manifold and b a fixed point of M. Then corresponding to each subgroup $H \subset \pi_1(M, b)$ there is a covering $F: \tilde{M} \to M$ such that for some $\tilde{b} \in F^{-1}(b)$ we have $F_* \pi_1(\tilde{M}, \tilde{b}) = H$. F and $\tilde{M}$ are unique to within isomorphism.*

Proof The uniqueness is just the previous theorem, and the proof of that theorem also indicates how the space must be constructed. The points of $\tilde{M}$ will consist of equivalence classes of paths from b, two such paths f, g being equivalent if and only if $f(1) = g(1)$ and $[f * g^{-1}] \in H$, g^{-1} denoting the path $g^{-1}(s) = g(1 - s)$, $0 \leq s \leq 1$. It follows from the fact that H is a subgroup that this is an equivalence; we denote it by $f \approx g$ and denote by $\{f\}$ the equivalence class of f (or point of $\tilde{M}$). The projection map $F: \tilde{M} \to M$ is defined by $F(\{f\}) = f(1)$ for any $f \in \{f\}$. Given any $\{f\} \in \tilde{M}$, let $p = f(1)$ and V, ψ be a coordinate neighborhood of p on M with $\psi(p) = 0$ and $\psi(V) = B_1^n(0)$, the open n-ball. For each $q \in V$ there is a unique path g_q from p to q corresponding to a radial line in $\psi(V)$. Then $q \to \{f * g_q\}$ defines a map $\theta_f: V \to \tilde{M}$ with $F \circ \theta_f(q) = F\{f * g_q\} = f * g_q(1) = q$ for all q in V. Suppose h is a path from b to p also and that $h \not\approx f$, that is, $[h \circ f^{-1}] \notin H$. Then it is easy to see that $\theta_f(V) \cap \theta_h(V) = \varnothing$. Indeed, if for some $q \in V$, we have $\{f * g_q\} = \{h * g_q\}$; this would require $[f * g_q * (h * g_q^{-1})] = [f * h^{-1}]$ to be an element of H, contrary to assumption. We leave it to the reader to check that the sets $\theta_f(V)$ with coordinate maps $\psi \circ F$ define a manifold[†] structure on $\tilde{M}$ which makes $F: \tilde{M} \to M$ a covering with $\{V, \psi\}$ as admissible neighborhoods.

Finally, we must establish that $F_*(\pi_1(\tilde{M}, \tilde{b})) = H$, where $\tilde{b} = \{e_b\}$, the point of $\tilde{M}$ determined by the constant path at b. Suppose that $f(t)$, $0 \leq t \leq 1$, is a loop at b with $[f] \in H$. Then $f(0) = f(1) = b$ and we define a one-parameter family f_t of paths from b by $f_t(s) = f(st)$, $0 \leq s, t \leq 1$. Let $\tilde{f}(t) = \{f_t(s)\}$. Then $\tilde{f}(t)$, $0 \leq t \leq 1$, is a path on $\tilde{M}$ with $F(\tilde{f}(t)) = f_t(1) = f(t)$, hence $\tilde{f}$ covers f and is a loop at $\tilde{b}$. It is straightforward to check that this actually determines an isomorphism F_* of $\pi_1(\tilde{M}, \tilde{b})$ onto H; we may apply methods similar to those already used above. This completes the proof. ∎

If we take $H = \{1\}$ we have a very important corollary.

[†] It is not easy to show that $\tilde{M}$ has a countable basis of open sets. A direct proof may be avoided; lift a Riemannian metric from M to $\tilde{M}$ (Exercise 1) and use Kobayashi and Nomizu [1, Vol. I, App. 2].

(9.8) Corollary *Every connected manifold M has a simply connected covering—called its inverse covering—which is unique to within isomorphism. Choice of $\tilde{b} \in F^{-1}(b)$ for $b \in M$ determines an isomorphism of $\pi_1(M, b)$ onto $\tilde{\Gamma}$ the group of covering transformations. Then $\tilde{M}/\tilde{\Gamma}$ is diffeomorphic to M, that is, M is the orbit space of its fundamental group acting properly discontinuously on its universal covering $\tilde{M}$.*

Exercises

1. Show that if $F\colon \tilde{M} \to M$ is a covering and M is Riemannian, there is a unique Riemannian metric on $\tilde{M}$ such that F is a local isometry.
2. Suppose that the assumptions of Exercise 1 are satisfied and that $\tilde{M}$ is compact. Determine the relation of the volumes of $\tilde{M}$ and M in terms of their fundamental groups.
3. Determine the meaning, in terms of $F\colon \tilde{M} \to M$ (a covering), of $F_*(\pi_1(\tilde{M}, \tilde{b}))$ being a *normal* subgroup of $\pi_1(M, F(\tilde{b}))$.
4. The universal covering $\tilde{G}$ of a connected Lie group G has a Lie group structure such that the covering map $F\colon \tilde{G} \to G$ is a homomorphism.

Notes

Integration on manifolds has two very important applications both of which are introduced briefly in this chapter. First, integration on Lie groups with respect to an invariant or bi-invariant volume element has been crucial in many areas of research on Lie groups and their homogeneous spaces. Weyl used it to prove the complete reducibility of representations of semisimple Lie groups, a central fact of representation theory (see, for example, Weyl [2]). We have seen how this was done for the compact case; the generalization to noncompact, semisimple groups was accomplished by Weyl's "unitary trick." But the integral is also used extensively to study function spaces on Lie groups and to prove such basic theorems as the Peter–Weyl theorem. The reader wishing to go further into these ideas should read relevant portions of Chevalley [1] and look at the last chapter of Helgason [1]. Many ideas used in the study of Lie groups can be exploited in studying the spaces on which they act, especially homogeneous spaces—which are essentially coset spaces of Lie groups. In this case, too, the invariant volume element and related theory of integration is crucial in many problems of current interest in analysis. Again the reader is referred to Helgason [1] for samples of these applications.

The other important application of integration on manifolds is to algebraic topology via de Rham's theorem and Hodge's theorem (see Warner [1]). This approach to topology was particularly useful in the study of the topology of Lie groups and homogeneous spaces, for which purpose de Rham's theorem was presumably first conjectured. A survey article by Samelson [1] should give the reader some idea of just how crucial this was in the early theory. A well-known theorem, the Gauss–Bonnet theorem (O'Neill [1]; Stoker [1]), is a beautiful example of how integration may be used to obtain relations between the topology of a manifold and some of its local geometric invariants: in this case the curvature. A recent treatise by Greub *et al.* [1] gives a comprehensive treatment of the many relations of differential geometry to algebraic topology, including, of course, the generalized Gauss–Bonnet theorem.

Both Milnor [2] and Guillemin and Pollack [1] have many elegant applications of differentiable methods to algebraic topology which are within the grasp of the reader at this point: determination of $H^*(S^n)$, proof of the fundamental theorem of algebra by use of the notion of index of a vector field, and the Poincaré-Hopf theorem on the sum of indices of singularities of a vector field.

VII DIFFERENTIATION ON RIEMANNIAN MANIFOLDS

We begin this chapter by showing very briefly how differential calculus can be applied to study the geometry of curves in Euclidean space E^n (or R^n), especially plane curves ($n = 2$) and space curves ($n = 3$). The geometric concepts which are discussed—arclength, curvature, and torsion—will be familiar to many readers as will the basic tool used, namely differentiation of a vector field along a curve. In the second section differentiation of vector fields along curves is used again to define and study differentiation of vector fields on a special class of Riemannian manifolds—those which are imbedded (or immersed) in E^n and carry the induced Riemannian metric. These same ideas can be used to investigate the geometry of surfaces in E^3, which is the subject matter of much of classical differential geometry. However, our main objective is to use this situation as a model in order to define differentiation of vector fields on an arbitrary Riemannian manifold M. This is done in Section 3 where the Riemannian connection ∇ is defined and its existence and uniqueness (depending only on the Riemannian metric) is demonstrated. We define in essence, a sort of directional derivative of vector fields Y on M, $\nabla_{X_p} Y$ giving the rate of change of Y at $p \in M$ in the direction of X_p. It generalizes $X_p f$ the derivative of a function, which was defined at the beginning of Chapter IV. As might be expected, it is more complicated than differentiation in Euclidean space, where we can take advantage of the natural parallelism. Conversely, it can itself be used to introduce a more restricted type of parallelism on arbitrary Riemannian manifolds. Once the basic properties of $\nabla_X Y$ are established, we are ready to apply differential calculus to the study of Riemannian manifolds.

We very briefly and formally introduce the Riemannian curvature tensor—an important geometric object which we study in the next chapter. The remainder of this chapter is spent on

the study of geodesics, which generalize to Riemannian manifolds the straight lines of Euclidean geometry. They have important similarities to straight lines as well as important differences. For example, like straight lines of Euclidean space, the unit tangent vector is constant (has derivative zero) as we move along the curve. Although geodesics in general are not the curves of shortest length between *any* two of their points, they will have this property for nearby points, and so on. Only the most basic properties can be proved in such a brief treatment, but enough is established to reveal the interesting variety of phenomena which occur for geodesics in general manifolds.

In the final section some important examples are considered, namely the Riemannian symmetric spaces, which have even more similarities with Euclidean space than the usual Riemannian manifold. In particular, the space of non-Euclidean geometry is a symmetric space. The examples considered here also have important curvature properties which are discussed (in part) in Chapter VIII.

The presentation of Sections 7 and 8 was very much influenced by Milnor [1], especially his Sections 10, 20, and 21.

1 Differentiation of Vector Fields along Curves in R^n

In order to clarify the definition of differentiation given later for general Riemannian manifolds, we shall first consider some special situations in the oriented Riemannian manifold R^n (with the standard orientation and inner product). In doing so we will make full use of the natural parallelism in R^n, that is, the natural identification of the tangent spaces at distinct points.

Let C be a curve in R^n given by $x(t) = (x^1(t), \ldots, x^n(t))$ with $a < t < b$. We suppose that $Z(t) = Z_{x(t)}$ is a vector field defined along C; thus to each $t \in (a, b)$ is assigned a vector (Fig. VII.1):

$$Z(t) = \sum a^i(t)\left(\frac{\partial}{\partial x^i}\right)_{x(t)} \in T_{x(t)}(R^n).$$

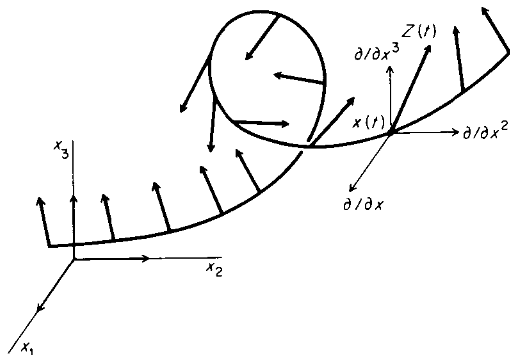


Figure VII.1

We will suppose Z to be of class C^1, at least, which means that the components $a^i(t)$ are continuously differentiable functions of t on the interval (a, b). The velocity vector of the (parametrized) curve itself is an example—in this case $a^i(t) = \dot{x}^i(t)$.

Our purpose is to define a derivative or rate of change of $Z(t)$ with respect to t; it will be denoted $\dot{Z}(t)$ or dZ/dt and will again be a vector field along the curve. Of course, in general, neither $Z(t)$ nor its derivative need be tangent to the curve. Now since in $\boldsymbol{R}^n$ we have a natural parallelism (or natural isomorphism) of $T_p(\boldsymbol{R}^n)$ and $T_q(\boldsymbol{R}^n)$ for any distinct $p, q \in \boldsymbol{R}^n$, we are able to give meaning to $Z(t_0 + \Delta t) - Z(t_0)$, the difference of a vector in $T_{x(t_0+\Delta t)}(\boldsymbol{R}^n)$ and a vector in $T_{x(t_0)}(\boldsymbol{R}^n)$. For definiteness we suppose $Z(t_0 + \Delta t)$ moved to or identified with the corresponding vector in $T_{x(t_0)}(\boldsymbol{R}^n)$ and that the subtraction is performed there (Fig. VII.2). This allows us to define the differential quotient

$$\frac{1}{\Delta t}[Z(t_0 + \Delta t) - Z(t_0)] = \sum_{i=1}^{n} \frac{a^i(t_0 + \Delta t) - a^i(t_0)}{\Delta t}\left(\frac{\partial}{\partial x^i}\right)_{x(t_0)}$$

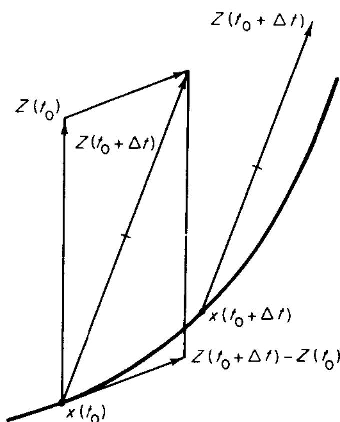


Figure VII.2

The equality is due to the fact that if we write vectors in terms of the basis $\partial/\partial x^1, \ldots, \partial/\partial x^n$ which is a field of *parallel* frames on $\boldsymbol{R}^n$, then vectors at distinct points, say $Z(t_0 + \Delta t)$ and $Z(t_0)$, are parallel if and only if they have the same components. Passing to the limit as $\Delta t \to 0$ gives the definition

(1.1) $$\dot{Z}(t_0) = \left(\frac{dZ}{dt}\right)_{t_0} = \sum \dot{a}^i(t_0)\left(\frac{\partial}{\partial x^i}\right)_{x(t_0)} \in T_{x(t_0)}(\boldsymbol{R}^n).$$

(1.2) Remark A consequence of this formula is that if we introduce a new parameter on the curve, say s, by $t = f(s)$ with $t_0 = f(s_0)$, then

$$\left(\frac{dZ}{ds}\right)_{s_0} = \left(\frac{dt}{ds}\right)_{s_0}\left(\frac{dZ}{dt}\right)_{t_0};$$

$(dt/ds)_{s_0}$ is a scalar; the other terms are vectors.

As a simple example consider the curve $x(t) = (\cos t, \sin t)$, a unit circle in $\boldsymbol{R}^2$. Suppose $Z(t) = -\sin t(\partial/\partial x) + \cos t(\partial/\partial y)$; this is the velocity vector of the point which traces out the circle. Then $dZ/dt = -\cos t(\partial/\partial x) - \sin t(\partial/\partial y)$ is a vector at $x(t) = (\cos t, \sin t)$ which has constant length $+1$ and points toward the origin (Fig. VII.3).

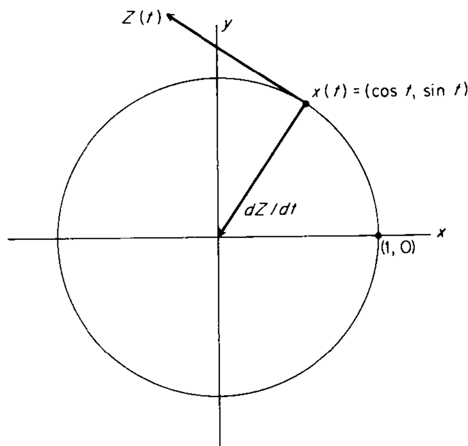


Figure VII.3

(1.3) Definition A vector field $Z(t)$ is *constant* or *parallel* along the curve $x(t)$ if and only if $dZ/dt = 0$ for all t.

Suppose that $Z_1(t)$ and $Z_2(t)$ are vector fields of the above type defined along the same curve C and that $f(t)$ is a differentiable function of t on $a < t < b$. Then $f(t)Z(t)$ and $Z_1(t) + Z_2(t)$ are vector fields along C and we have the following easy consequences of formula (1.1).

(1.4a) $$\frac{d}{dt}(Z_1(t) + Z_2(t)) = \frac{dZ_1}{dt} + \frac{dZ_2}{dt},$$

(1.4b) $$\frac{d}{dt}(f(t)Z(t)) = \frac{df}{dt}Z(t) + f(t)\frac{dZ}{dt},$$

(1.4c) $$\frac{d}{dt}(Z_1(t), Z_2(t)) = \left(\frac{dZ_1}{dt}, Z_2(t)\right) + \left(Z_1(t), \frac{dZ_2}{dt}\right),$$

where (Z_1, Z_2) is the standard inner product in $\boldsymbol{R}^n$.

The formula we have given for dZ/dt is in terms of the components of $Z(t)$ relative to the natural field of frames $\partial/\partial x^1, \ldots, \partial/\partial x^n$ in $\boldsymbol{R}^n$, which are constant along $x(t)$. However, we sometimes find it convenient to use some other field of frames, say $F_1(t), \ldots, F_n(t)$, defined and of class C^1 at least

along $x(t)$. Then $Z(t)$ has a unique expression as a linear combination of these vectors at each $x(t)$:

$$Z(t) = b^1(t)F_1(t) + \cdots + b^n(t)F_n(t).$$

Differentiating this expression we obtain, with the aid of (1.4a,b),

$$\frac{dZ}{dt} = \sum_{j=1}^{n} \left(\frac{db^j}{dt} F_j(t) + b^j(t) \frac{dF_j}{dt} \right).$$

However, since dF_j/dt are vectors along $x(t)$, they too are linear combinations of $F_k(t)$,

$$\frac{dF_j}{dt} = \sum_{k=1}^{n} a_j^k(t) F_k(t).$$

This gives the formula

$$\text{(1.5)} \qquad \frac{dZ}{dt} = \sum_k \left(\frac{db^k}{dt} + \sum_j b^j(t) a_j^k(t) \right) F_k(t).$$

This includes (1.1) as a special case since $a_j^k(t) \equiv 0$ when the frames $F_1(t), \ldots, F_n(t)$ are parallel.

Although we have used a particular coordinate system, the natural one in $\boldsymbol{R}^n$, in fact

$$\frac{dZ}{dt} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [Z(t_0 + \Delta t) - Z(t_0)]$$

is an object which depends only on the geometry of the space; it is independent of coordinates, so it is defined equally well for parametrized curves in $\boldsymbol{E}^n$.

The Geometry of Space Curves

As an illustration of these ideas we derive the *Frenet–Serret formulas* for a curve of class C^3 in $\boldsymbol{R}^3$. We first note that length of the curve from a fixed point $x_0 = x(t_0)$ is given by $s = \int_{t_0}^{t} ((\dot{x}(t), \dot{x}(t))^{1/2}\, dt$ so that $ds/dt = (\dot{x}(t), \dot{x}(t))^{1/2}$. If s is used as parameter, then $ds/dt \equiv ds/ds \equiv 1$ so that $\dot{x}(s)$ is a unit vector tangent to the curve. We let $T(s) = \dot{x}(s)$ denote this unit tangent vector. Because arclength, the parameter s (to within an additive constant), and $T(s)$ are determined by the (induced) Riemannian metric on $x(s)$, not by the particular rectangular Cartesian coordinates or origin used, they and the derivatives of $T(s)$ are geometric invariants of the curve, that is, they will be the same at corresponding points for congruent curves. Differentiating the identity $(T(s), T(s)) \equiv 1$ and using (1.4c), we obtain $2(T(s), dT/ds) \equiv 0$.

Therefore dT/ds is zero, or is a nonzero vector orthogonal to $T(s)$ at each point of the curve. We define the *curvature* $k(s)$ by $k(s) = \|dT/ds\|$; and when $k(s) \neq 0$, we let $N(s)$ be the unique unit vector defined by $dT/ds = k(s)N(s)$. When $k(s) \neq 0$, let $B(s)$ be the uniquely determined unit vector such that $T(s), N(s), B(s)$ define an orthonormal frame with the orientation of $\partial/\partial x^1$, $\partial/\partial x^2$, $\partial/\partial x^3$ (see Fig. VII.4). This determines a field of orthonormal frames

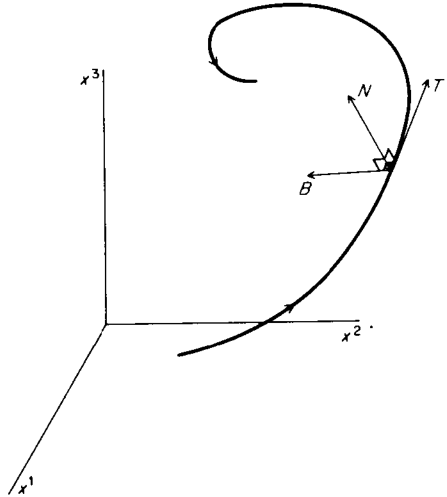


Figure VII.4

along the curve, defined whenever $k(s) \neq 0$, which we shall henceforth assume is the case at all points of the curve under consideration. This assumption is justified since it is the generic or typical situation for a space curve. In fact, we have the following fact concerning a curve for which $k(s)$ vanishes identically.

(1.6) Theorem *If $k(s) \equiv 0$ on the interval of definition, then $x(s)$ is a straight line segment on that interval, and conversely, for a straight line, $k(s) \equiv 0$.*

Proof If the curve is a straight line, then it is given in terms of arclength by $x^i(s) = a^i + b^i s$, $i = 1, 2, 3$, where $\sum_{i=1}^3 (b^k)^2 = 1$. Thus

$$T = \sum_{i=1}^3 b^i \, \partial/\partial x^i$$

and $dT/ds \equiv 0$. Conversely, if $k(s) \equiv 0$, then $dT/ds \equiv 0$. Since $T = \sum dx^i/ds \; \partial/\partial x^i$, s being arclength, this implies $d^2x^i/ds^2 \equiv 0$, $i \equiv 1, 2, 3$. Thus $x^i(s) = a^i + b^i s$, $i = 1, 2, 3$, with a^i and b^i constants, and the curve is a straight line. Note that $T(s)$ and $k(s)$ are defined for a curve in $\boldsymbol{R}^n$, any n (not just $n = 3$), and the proposition just proved is still valid. ∎

Now, returning to the study of a curve in three-space, for convenience of notation we let $F_1(s)$, $F_2(s)$, $F_3(s)$ denote $T(s)$, $N(s)$, $B(s)$, respectively. Since

this is a field of orthonormal frames we have $(F_i(s), F_j(s)) \equiv \delta_{ij}$. Differentiation of these equations gives the relations

$$\left(\frac{dF_i}{ds}, F_j(s)\right) + \left(F_i(s), \frac{dF_j}{ds}\right) \equiv 0, \qquad i, j = 1, 2, 3.$$

As we pointed out in the derivation, dF_j/ds must be a linear combination of the $F_k(s)$ for every s, so we may write

$$\frac{dF_j}{ds} = \sum_k a_j^k F_k(s), \qquad j = 1, 2, 3.$$

This combines with the previous equations to give

$$\left(\sum_k a_i^k F_k, F_j\right) + \left(F_i, \sum_k a_j^k F_k\right) \equiv 0$$

or

$$a_i^j(s) + a_j^i(s) \equiv 0, \qquad 1 \le i, j \le 3.$$

This means that the matrix $(a_j^i(s))$ is skew-symmetric. By definition $dT/ds = k(s)N$, that is, $a_1^2(s) = k(s)$, and so $a_3^1(s) \equiv 0 \equiv a_1^3(s)$. Let $a_2^3(s) = \tau(s)$ as a matter of notation. Then rewriting in terms of T, N, B, we have the Frenet–Serret formulas:

(1.7)
$$\begin{aligned}
\frac{dT}{ds} &= \qquad\quad k(s)N, \\
\frac{dN}{ds} &= -k(s)T \qquad\qquad + \tau(s)B, \\
\frac{dB}{ds} &= \qquad -\tau(s)N,
\end{aligned}$$

expressing the derivatives with respect to s of T, N, and B which are called the *tangent*, *normal*, and *binormal* vectors, respectively, of $x(s)$, in terms of these vectors themselves. Formula (1.7) defines the functions $k(s)$ and $\tau(s)$ along the curve.

(1.8) Definition $k(s)$ is called the *curvature* and $\tau(s)$ the *torsion* of the curve C at $x(s)$.

The curvature measures deviation of C from a straight line and the torsion measures "twisting" or deviation of C from being a plane curve. Of course, T, N, and B as well as curvature and torsion are independent of the coordinates used in the Euclidean space containing C.

(1.9) Theorem *A curve in E^3 lies in a plane if and only if $\tau(s) \equiv 0$.*

Proof If the curve lies in a plane, then from the definition of $T(s)$ and dT/ds we see that these vectors lie in the plane of the curve for each point $x(s)$ of the curve. Thus the unit vector $B(s)$ has a fixed direction, orthogonal

to the plane, and so is always parallel to a fixed unit vector orthogonal to the plane. Therefore $dB/ds \equiv 0$ and $\tau(s) \equiv 0$.

Suppose that $\tau(s) \equiv 0$. Then $dB/ds \equiv 0$ and B is a constant vector along the curve. We choose the coordinate axes so that the curve passes through the origin 0 at $s = 0$ and so that $B(s)$ is parallel to $\partial/\partial x^3$, the unit vector in the direction of the x^3-axis. Then $x(s) = (x^1(s), x^2(s), x^3(s))$ determines the vector $\mathbf{x}(s)$ from the origin 0 to the point $x(s)$ on the curve. Differentiating $(\mathbf{x}(s), B(s))$, we have

$$\frac{d}{ds}(\mathbf{x}(s), B(s)) \equiv (T(s), B(s)) + \left(\mathbf{x}(s), \frac{dB}{ds}\right) = (T(s), B(s)) \equiv 0$$

so that $(x(s), B(s))$ is constant. Since $x(s_0) = 0$, that is, $\mathbf{x}(s_0) = 0$, the vector $\mathbf{x}(s)$ [or line $0x(s)$] is always perpendicular to $B = \partial/\partial x^3$. Thus the curve lies in the x^1x^2-plane. ∎

The advantage of using arclength s as parameter on the curve C and the frames $T(s)$, $N(s)$, $B(s)$ is that they all have intrinsic geometric meaning, depending as they do only on the Riemannian structure of the ambient space $\boldsymbol{E}^3$ and the nature of the curve itself and not on any coordinates that we might use in $\boldsymbol{E}^3$. Although the formulas for the derivative of a vector field Z along the curve are more complicated, since they involve the second terms a_j^k in (1.5) which vanish when we use parallel frames along C, nevertheless the advantage of being geometrically determined is quite crucial. Even the coefficients a_j^k, which here are $\pm k(s)$ and $\pm\tau(s)$ or zero, have geometric meaning as we have seen. In fact it is not difficult to show (see, for example, O'Neill [1]) that $k(s)$ and $\tau(s)$ determine C up to congruence.

As an illustration of the utility of the Frenet frames T, N, B we consider briefly the dynamics of a moving particle in space whose position $p(t)$ is given as a function of time t. Let $s(t)$ be the length of path traversed from time $t = 0$ to time t, $s(t) = \int_0^t ((dp/dt, dp/dt))^{1/2}\, dt$. Then $ds/dt = ((dp/dt, dp/dt))^{1/2} = \|dp/dt\|$ is the *speed* with which the particle moves along the curve, while its *velocity vector* is given by

$$\mathbf{v}(t) = \frac{dp}{dt} = \frac{dp}{ds}\frac{ds}{dt} = T\frac{ds}{dt},$$

where T is the unit tangent vector. Differentiating we obtain the acceleration

$$\mathbf{a}(t) = \frac{d^2p}{dt^2} = \frac{dT}{ds}\left(\frac{ds}{dt}\right)^2 + T\frac{d^2s}{dt^2}.$$

Since $dT/ds = kN$ this becomes

$$\textbf{(1.10)} \qquad \mathbf{a}(t) = \frac{d^2s}{dt^2}T + k\left(\frac{ds}{dt}\right)^2 N.$$

Thus the acceleration decomposes into the sum of two vectors, one in the direction of the curve, whose magnitude is the time rate of change of the speed d^2s/dt^2, and the other normal to the curve and directly proportional to both the square of the speed and to the curvature, this latter depending only on the curve. If the motion is a straight line motion, then $k = 0$ so that $\mathbf{a}$ has the direction of the line. If the particle moves at constant speed, so that $d^2s/dt^2 = 0$, then the acceleration depends only on the shape of the path. The same remarks also apply to the force F acting on the particle, which by Newton's second law, $F = m\mathbf{a}$, is proportional to $\mathbf{a}$ with the mass m as constant of proportionality.

Curvature of Plane Curves

Some special comment is required for the case of a curve C lying on an oriented plane. Let $s \to (x(s), y(s))$ define the curve, parametrized by arclength. Then $T = \dot{x}(s)\,\partial/\partial x + \dot{y}(s)\,\partial/\partial y$ is the unit tangent vector. If $dT/ds \neq 0$, then we may as before define $k(s) = \|dT/ds\|$, that is, we may consider the curve as a space curve $(x(s), y(s), 0)$ whose z-coordinate $z(s) \equiv 0$, and use the same definitions. However, for plane curves a more refined definition of curvature is possible: At each point of C choose N so that T, N have the same orientation as $\partial/\partial x, \partial/\partial y$ (this uniquely determines T, N). (See Fig. VII.5.) Then define the curvature $\tilde{k}(s)$ so that $\tilde{k}(s)N = dT/ds$. This allows $\tilde{k}(s)$ to be negative, zero, or positive. The curvature thus defined for a plane curve has the previously defined curvature of C (considered as a space curve) as its absolute value, $k(s) = |\tilde{k}(s)|$. To carry our interpretation somewhat further let $\theta(s)$ be the angle of T with the positive x-axis (Fig.VII.5). Then

$$T(s) = \cos\theta(s)\frac{\partial}{\partial x} + \sin\theta(s)\frac{\partial}{\partial y},$$

and

$$\frac{dT}{ds} = -\dot{\theta}(s)\sin\theta(s)\frac{\partial}{\partial x} + \dot{\theta}(s)\cos\theta(s)\frac{\partial}{\partial y}.$$

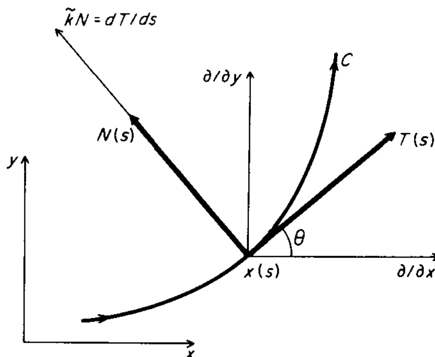


Figure VII.5

The unit vector $N(s)$ chosen so that $T(s)$, $N(s)$ is an oriented orthonormal basis is $N(s) = -\sin\theta(\partial/\partial x) + \cos\theta(\partial/\partial y)$, since the determinant of the coefficients of T, N as combinations of $\partial/\partial x$, $\partial/\partial y$ is

$$\det\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = +1.$$

Thus $\tilde{k}(s) = \dot{\theta}(s)$ or $d\theta/ds$, the rate of turning of the tangent vector T with respect to arclength. Moving along C in the direction of increasing s the curvature is positive when the tangent is turning counterclockwise and negative otherwise. Its sign depends on the *sense* of the curve (direction of increasing s) and the *orientation* of the plane, but not on the coordinates.

Suppose C is a circle of radius r. Then $s \to (r\cos(s/r), r\sin(s/r))$ gives the curve parametrized by arclength:

$$T = -\sin\left(\frac{s}{r}\right)\frac{\partial}{\partial x} + \cos\left(\frac{s}{r}\right)\frac{\partial}{\partial y},$$

$$\tilde{k}N = \frac{dT}{ds} = -\frac{1}{r}\cos\left(\frac{s}{r}\right)\frac{\partial}{\partial x} - \frac{1}{r}\sin\left(\frac{s}{r}\right)\frac{\partial}{\partial y}.$$

Since

$$N = -\cos\left(\frac{s}{r}\right)\frac{\partial}{\partial x} - \sin\left(\frac{s}{r}\right)\frac{\partial}{\partial y}$$

is the unique unit vector such that T, N has the orientation of $\partial/\partial x$, $\partial/\partial y$, we see that $\tilde{k}(s) = 1/r$. Thus the curvature is a constant. If, as we have assumed by our parametrization, the circle is traversed in the counterclockwise sense, it is a positive number; in any case its magnitude is inversely proportional to the radius.

Returning momentarily to the dynamics of a moving particle we see that if a particle moves on a circle in such a way that its speed is constant v_0, then the force F acting on the particle is

$$F = m\mathbf{a} = \frac{mv_0^2}{r}N.$$

Since N is the unit normal vector, F is directed toward the center of the circle and its magnitude is mv_0^2/r which gives the usual formula for the centripetal force necessary to keep the particle in a circular orbit.

Exercises

1. Prove that a curve in $\boldsymbol{R}^3$ for which $\tau(s) \equiv 0$ and $k(s)$ is a constant, $k \neq 0$, is a circle.

2. A helix is defined by parametric equations of the form $x^1 = a\cos t$, $x^2 = a\sin t$ and $x^3 = bt$, where a, b are positive constants. Determine k and τ for a helix. [*Hint:* First change to arclength as parameter.]
3. Let Z_1 and Z_2 be two vector fields along a curve in $\boldsymbol{R}^3$ and let $Z_1 \times Z_2$ be their cross product (as defined in a three-dimensional vector space). Show that

$$\frac{d}{dt}(Z_1 \times Z_2) = \frac{dZ_1}{dt} \times Z_2 + Z_1 \times \frac{dZ_2}{dt}.$$

4. Show that the plane which closest approximates a curve C at p is spanned by T and N. This plane is the limiting position of the plane through p', p, p'' on C as p', p'' approach p.
5. Using the technique of Exercise 4, find the center of the best approximating circle and of the best approximating sphere to each point p of C. [They should be located relative to the moving coordinate frame T, N, B.]
6. Show that if the plane of T and N goes through some fixed point 0 of $\boldsymbol{E}^3$ for every point on C, then C lies in a plane.

2 Differentiation of Vector Fields on Submanifolds of $\boldsymbol{R}^n$

In the previous section we studied differentiation of vector fields along curves, which includes, of course, one-dimensional submanifolds of Euclidean space. In this section we do the same for vector fields "along" other submanifolds $M \subset \boldsymbol{R}^n$, for example a surface in $\boldsymbol{R}^3$. This is somewhat more complicated and certainly not the most direct way of approaching the subject of differentiation on manifolds; but it should help our geometric understanding. Just as in the case of a curve, we are concerned with a vector field Z defined at each point of M but not necessarily tangent to M (see Fig.VII.6), that is, to each $p \in M$, we assign $Z_p \in T_p(\boldsymbol{R}^n)$. When Z is such that

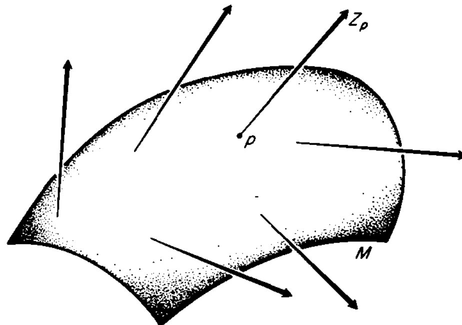


Figure VII.6

Z_p is tangent to M, $Z_p \in T_p(M) \subset T_p(\boldsymbol{R}^n)$, then we shall say that Z is a vector field *on* M or a *tangent* vector field. Only in this case has Z meaning for M as an abstract manifold, independent of any imbedding or immersion† in $\boldsymbol{R}^n$. In any case differentiability of Z may be given meaning since its components relative to the canonical frames of $\boldsymbol{R}^n$ at points of M will be functions on M, $Z_p = \sum_{\alpha=1}^{n} a^\alpha(p)(\partial/\partial x^\alpha)_p$, and by definition we shall say that Z is of *class* C^r if $a^\alpha(p)$, $\alpha = 1, \ldots, n$, are of class C^r on M. In particular, the vector fields $\partial/\partial x^1, \ldots, \partial/\partial x^n$ of $\boldsymbol{R}^n$, restricted to M, are C^∞-vector fields *along* M (but rarely *on* M).

If $p \in M$, then $T_p(\boldsymbol{R}^n)$ and its subspace $T_p(M)$ carry the standard inner product of $\boldsymbol{R}^n$ so M has the induced Riemannian metric. This allows us to decompose any vector Z_p, $p \in M$, in a unique way into $Z_p = Z'_p + Z''_p$ with $Z'_p \in T_p(M)$ and $Z''_p \in T_p^\perp(M)$, the orthogonal complement of $T_p(M)$. This reflects the direct sum decomposition of $T_p(\boldsymbol{R}^n)$ into mutually orthogonal subspaces: $T_p(\boldsymbol{R}^n) = T_p(M) \oplus T_p^\perp(M)$ called the *tangent* space and the *normal* space to M at p. Let π', π'' denote the corresponding projections: $\pi'(Z_p) = Z'_p$ and $\pi''(Z_p) = Z''_p$; they are linear mappings of $T_p(\boldsymbol{R}^n)$ onto the subspaces tangent and normal to M. Figures VII.7 and VII.8 illustrate this decomposition for a curve and surface in $\boldsymbol{R}^3$.

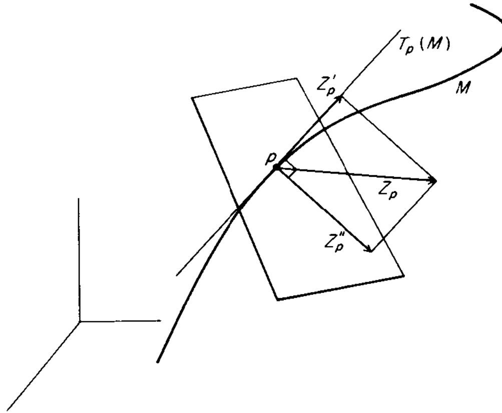


Figure VII.7

Suppose that Z is a vector field along M of class C^r. Then $\pi'(Z)$ and $\pi''(Z)$ are also vector fields, which are tangent and normal to M, provided that they are differentiable. We leave the proof of the following lemma (including the assertion of C^r differentiability) to the exercises.

(2.1) Lemma *With Z as above $\pi'(Z)$ and $\pi''(Z)$ define mutually orthogonal C^r-vector fields Z', Z'' along M such that $Z = Z' + Z''$, that is, at each $p \in M$,*

† Since we consider only local questions in this section, we may restrict ourselves to imbedded (regular) submanifolds by Theorem III.4.12.

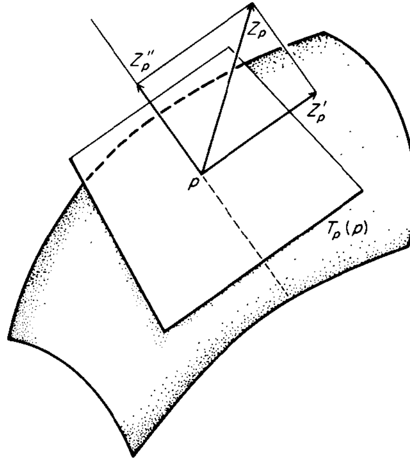


Figure VII.8

$Z'_p \in T_p(M)$ and $(Z'_p, Z''_p) = 0$. If f is a function of class C^r on M, then $\pi'(fZ) = f\pi'(Z)$ and $\pi''(fZ) = f\pi'(Z)$. Further, given two such vector fields Z_1, Z_2, then $\pi'(Z_1 + Z_2) = \pi'(Z_1) + \pi'(Z_2)$ and $\pi''(Z_1 + Z_2) = \pi''(Z_1) + \pi''(Z_2)$.

As examples we note that a vector field Z along a curve decomposes uniquely into the sum of a tangent vector field and a vector field in the normal plane: $\pi'(Z) = (Z, T)T$ and $\pi''(Z) = (Z, N)N + (Z, B)B$ (see Fig. VII.7). For the case of an arbitrary C^∞ imbedded manifold M, we see that $\pi'(\partial/\partial x^\alpha)$ applied at each $p \in M$ gives a C^∞ tangent vector field to M for each $\alpha = 1, \ldots, n$.

Now let Y be a tangent vector field to $M \subset \boldsymbol{R}^n$, that is, for each $p \in M$, $Y_p \in T_p(M)$, or equivalently $\pi'(Y) \equiv Y$. If $p(t)$ is a curve on M of class C^1 or higher, defined for some interval of values of t, then $Y(t) = Y_{p(t)}$ is a vector field along the curve. As such we can ignore M and differentiate $Y(t)$ as a vector field along a curve in $\boldsymbol{R}^n$ obtaining dY/dt, another vector field along the curve. In general, of course, dY/dt will not be tangent to M; however, at each point $p(t)$ we may project dY/dt to a tangent vector $\pi'(dY/dt)$.

(2.2) Definition The projection $\pi'(dY/dt)$ will be denoted DY/dt and will be called the *covariant derivative* of the tangent vector field Y on M along the curve $p(t)$ (see Fig. VII.9).

Both Y and DY/dt are tangent vector fields, and thus have meaning for the abstract manifold M. However, the process by which DY/dt is obtained from Y and $p(t)$ makes use of the imbedding of M in $\boldsymbol{R}^n$. Our ultimate aim is to obtain for an abstract Riemannian manifold M a definition of derivative

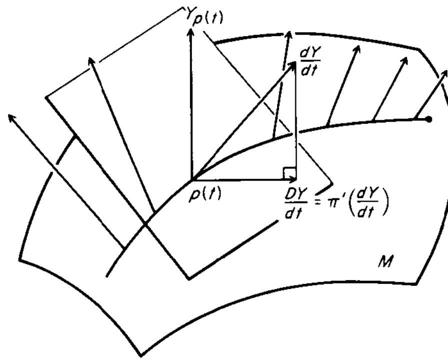


Figure VII.9

of a vector field along a curve which does not make use of an imbedding but is intrinsic to the Riemannian and differentiable structure of M itself. In the meantime we shall study in more detail the properties of DY/dt and will see that it has properties similar to the derivative discussed in Section 1.

It is important to note that $Y(t)$ need not be the restriction to a curve $p(t)$ of a vector field Y on M—as above—in order for DY/dt to be defined. We need only suppose that $Y(t)$ is a vector field along $p(t)$, so defined that it is always tangent to M, that is, such that $Y(t) \in T_{p(t)}(M)$. Then, as above, $DY/dt = \pi'(dY/dt)$, where dY/dt is the derivative of the vector field along a curve as defined in the previous section. Now suppose as in (1.4a–c) that we have vector fields $Y_1(t)$ and $Y_2(t)$ along $p(t)$ on M and tangent to M. Then we have corresponding properties.

(2.3) Theorem *With $Y(t)$, $Y_1(t)$, $Y_2(t)$ as above and $f(t)$ a C^1 function of t we have*

(2.3a) $$\frac{D}{dt}(Y_1 + Y_2) = \frac{DY_1}{dt} + \frac{DY_2}{dt},$$

(2.3b) $$\frac{D}{dt}(f(t)Y(t)) = \frac{df}{dt}Y(t) + f(t)\frac{DY}{dt},$$

(2.3c) $$\frac{d}{dt}(Y_1, Y_2) = \left(\frac{DY_1}{dt}, Y_2\right) + \left(Y_1, \frac{DY_2}{dt}\right).$$

The last equation concerns the induced Riemannian metric on M, that is, the inner product on $T_p(M)$, at each $p \in M$, induced by the inner product in $T_p(\boldsymbol{R}^n)$. These properties are immediate consequences of the definitions, the

properties of π', and of the corresponding statements of (1.4a–c). Applying π' to both sides of (1.4a) and using linearity gives (2.3a). Similarly using (1.4b), we see that (2.3b) holds:

$$\frac{D}{dt}(fY) = \pi' \frac{d}{dt}(fY) = \pi'\left(\frac{df}{dt} Y + f \frac{dY}{dt}\right) = \frac{df}{dt} Y + f \frac{DY}{dt}.$$

The last property follows from (1.4c) if we remark that

$$\frac{dY_i}{dt} = \pi'\left(\frac{dY_i}{dt}\right) + \pi''\left(\frac{dY_i}{dt}\right) = \frac{DY_i}{dt} + \pi''\left(\frac{dY_i}{dt}\right), \qquad i = 1, 2,$$

and that $\pi''(dY_i/dt)$ is orthogonal to $T_{p(t)}(M)$ so that

$$\left(\frac{DY_1}{dt} + \pi''\left(\frac{dY_1}{dt}\right), Y_2\right) + \left(Y_1, \frac{DY_2}{dt} + \pi''\left(\frac{dY_2}{dt}\right)\right) = \left(\frac{DY_1}{dt}, Y_2\right) + \left(Y_1, \frac{DY_2}{dt}\right).$$

(2.4) Remark If we change to a new parameter, say s, by $t = f(s)$, then $DY/ds = (DY/dt)(dt/ds)$, $dt/ds = f'(s)$ being a scalar. This also follows from applying π' to the similar relation $dY/ds = (dY/dt)(dt/ds)$ of the previous section.

(2.5) Definition Given $M \subset \boldsymbol{R}^n$ as above, let $Y_{p(t)}$ be a vector field defined at each point of a curve $p(t)$ on M and which at each point is tangent to M, that is, a vector field along $p(t)$ tangent to M. Then we shall say that $Y_{p(t)}$ is a *constant* or *parallel* vector field if $DY/dt \equiv 0$. More generally if Y is a tangent vector field on all of M, then it is *constant* or *parallel* if it has this property along every curve on M.

It is very important to note that DY/dt may be identically zero even though dY/dt is not, thus a vector field along a curve may be constant considered as a vector field on a submanifold M of $\boldsymbol{R}^n$ even though it is not constant considered as a vector field along the same curve in $\boldsymbol{R}^n$. This observation is a crucial point in some of our subsequent discussions, so we give a simple example. Let $M = S^1$, the unit circle in $\boldsymbol{R}^2$. Then $t \to (\cos t, \sin t)$ is the parametric representation and it may be considered as defining a curve on M. Let $Y(t)$ be the unit tangent vector to this curve. As we have seen dY/dt is orthogonal to $Y(t)$, that is, normal to M; hence $DY/dt = \pi'(dY/dt) \equiv 0$, although dY/dt is never zero and in fact has constant length $+1$. Since any great circle on the unit sphere $S^{n-1} \subset \boldsymbol{R}^n$ is congruent to the great circle $t \to p(t) = (\cos t, \sin t, 0, \ldots, 0)$ on the intersection of S^{n-1} and the 2-plane $x^3 = \cdots = x^n = 0$ of $\boldsymbol{R}^n$, we see that the unit tangent vector to any great circle arc $p(t)$, parametrized by arclength, has the same property:

$$\frac{DY}{dt} = \frac{D}{dt}\left(\frac{dp}{dt}\right) \equiv 0.$$

This is admittedly a special case; in general the derivative of a tangent vector field to M along a curve $p(t)$ in M has both normal and tangential components different from zero. If a curve on M is such that $(D/dt)\cdot(dp/dt) \equiv 0$, that is, the (covariant) derivative of the unit tangent vector to the curve is zero along the curve, then we shall say the curve is a *geodesic* of M. We have just seen that the great circles on the unit sphere in $\boldsymbol{R}^n$ are geodesics. In the case in which M is an open subset of $\boldsymbol{R}^n$ or all of $\boldsymbol{R}^n$, then $dY/dt = DY/dt$, that is, in $\boldsymbol{R}^n$ itself, as might be expected, covariant differentiation is just the usual differentiation. In this special case, according to Theorem 1.6, the only curves $p(t)$ for which

$$\frac{D}{dt}\left(\frac{dp}{dt}\right) = \frac{d}{dt}\left(\frac{dp}{dt}\right)$$

vanishes identically are straight lines parametrized by arclength—or with t proportional to arclength. Thus geodesics on an imbedded manifold M are those curves which in some sense generalize the concept of straight line—even though they may not look "straight" when viewed from the ambient space $\boldsymbol{R}^n$. We shall study these questions in some detail later in this chapter.

Formulas for Covariant Derivatives

In order to further study the covariant derivative DY/dt of a tangent vector field Y on M along a curve we will need more detailed computations using local coordinates. Suppose $\dim M = m$ and that U, φ is a local coordinate system on M with $\varphi(U) = W$, an open subset of $\boldsymbol{R}^m$. We denote the local coordinates by $u^1, \ldots, u^m$ and remark that $\varphi^{-1}\colon W \to R^n$ is an imbedding of W whose image is, of course, U—an open subset of M. We have previously referred to φ^{-1} as a *parametrization* of M. Let $u = (u^1, \ldots, u^m)$, then

$$\varphi^{-1}(u) = (g^1(u), \ldots, g^n(u)), \qquad u \in W,$$

gives φ^{-1} in terms of its coordinate mappings $g^\alpha(u)$. [We let α, β, γ, and so on, denote indices that range from 1 to n and i, j, k, and so on, indices ranging from 1 to m.] The coordinate frames will be denoted $F_1, \ldots, F_m$; they span the tangent space to M at each point. Since this tangent space $T_p(M)$ at $p \in M$ is a subspace of $T_p(\boldsymbol{R}^n)$, these vectors are linear combinations of $\partial/\partial x^1, \ldots, \partial/\partial x^n$. In fact, generalizing earlier formulas for $m = 2$ and $n = 3$ (Example IV.1.10) we have:

$$\textbf{(2.6)} \qquad F_{ip} = \varphi_*^{-1}\left(\frac{\partial}{\partial u^i}\right) = \sum_{\alpha=1}^{n} \left(\frac{\partial g^\alpha}{\partial u^i}\right)_{\varphi(p)} \frac{\partial}{\partial x^\alpha}.$$

Now suppose that $p(t)$ is a curve on M of class C^1 and that $Y(t) = Y_{p(t)}$ is a vector field along the curve which is always tangent to M. Then $Y(t)$ may

be written as a linear combination of $F_1, \ldots, F_m$, so that $Y(t) = \sum_{k=1}^m b^k(t)F_k$. Although

$$\frac{dY}{dt} = \sum \left(\frac{db^k}{dt} F_k + b^k \frac{dF_k}{dt}\right)$$

is not tangent to M in general, projecting we obtain

$$\frac{DY}{dt} = \pi'\left(\frac{dY}{dt}\right) = \sum_{k=1}^m \left(\frac{db^k}{dt} F_k + b^k \pi'\left(\frac{dF_k}{dt}\right)\right)$$

or

(2.7)
$$\frac{DY}{dt} = \sum_{k=1}^m \left(\frac{db^k}{dt} F_k + b^k \frac{DF_k}{dt}\right).$$

We know, however, that DF_i/dt, $i = 1, \ldots, m$, are vectors tangent to M and may be expressed as linear combinations of $F_1, \ldots, F_m$. Suppose that the curve $p(t)$ is given in local coordinates by $\varphi(p(t)) = (u^1(t), \ldots, u^m(t))$, then in (2.6) the components are (composite) functions $(\partial g^\alpha/\partial u^i)_{\varphi(p(t))}$ of t through $u^1(t), \ldots, u^m(t)$, and at each $p(t)$

$$\frac{DF_i}{dt} = \pi'\left(\frac{dF_i}{dt}\right) = \sum_{\alpha=1}^n \sum_{j=1}^m \frac{\partial g^\alpha}{\partial u^j \partial u^i} \frac{du^j}{dt} \pi'\left(\frac{\partial}{\partial x^\alpha}\right)$$

by the ordinary chain rule of differentiation applied to (2.6), and the properties of π'. The derivatives $\partial^2 g^\alpha/\partial u^j \, \partial u^i$ are functions of $u^1, \ldots, u^m$ and are evaluated at $u(t) = (u^1(t), \ldots, u^m(t))$ in this formula.

We have previously remarked that when M is imbedded in $\boldsymbol{R}^n$ by a C^∞ imbedding—which we shall always assume—then $\partial/\partial x^\alpha$ restricted to M is a C^∞ vector field along M. By Lemma 2.1, $\pi'(\partial/\partial x^\alpha)$ defines a C^∞ tangent vector field on M, which must have then a unique expression of the form $\pi'(\partial/\partial x^\alpha) = \sum_{k=1}^m a_\alpha^k(u)F_k$ on U. The $a_\alpha^k(u)$ are C^∞ functions on M which we do not compute. Using them and the coordinate functions $g^\alpha(u)$ of the parametrization φ^{-1} we define the C^∞ functions $\Gamma_{ij}^k(u)$ as

$$\Gamma_{ij}^k = \sum_\alpha \frac{\partial^2 g^\alpha}{\partial u^i \, \partial u^j} a_\alpha^k = \Gamma_{ji}^k, \qquad 1 \leq i, j, k \leq m.$$

Symmetry in i, j is a consequence of interchangeability of the order of differentiation. We do not explicitly compute these functions now, but we use them to write new formulas for DF_i/dt:

(2.8)
$$\frac{DF_i}{dt} = \sum_{j,k=1}^m \Gamma_{ij}^k \frac{du^j}{dt} F_k, \qquad i = 1, \ldots, m,$$

at each $p = p(t)$, the Γ_{ij}^k being evaluated at $(u^1(t), \ldots, u^m(t))$. A particular case, the curve given by $u^i =$ constant for $i \neq j$ and $u^j = t$, gives the formula

for the covariant derivative of the vector field F_i along the jth coordinate curve, conveniently denoted $DF_i/\partial u^j$:

$$\frac{DF_i}{\partial u^j} = \sum_k \Gamma_{ij}^k F_k .$$

This gives an interpretation of the meaning of $\Gamma_{ij}^k(u)$; it is the kth component (relative to the coordinate frames) of the covariant derivative of F_i along that curve in which only the jth coordinate is allowed to vary, that is, along a coordinate curve. Using these formulas we may finally write (2.7) in the form in which we want it (after appropriate change of indices):

$$\textbf{(2.9)} \qquad \frac{DY}{dt} = \sum_{k=1}^{m} \left(\frac{db^k}{dt} + \sum_{i,\, j=1}^{m} \Gamma_{ij}^k(u(t)) b^i(t) \frac{du^j}{dt} \right) F_k .$$

This formula gives us the analog to formula (1.5) in that DY/dt is expressed in terms of the field of frames $F_1, \ldots, F_s$ on $U \subset M$, frames defined independently of either $p(t)$ or Y. The components of the covariant derivative are the terms in brackets. The functions $\Gamma_{ij}^k(u)$ are defined over all of U and in (2.9) are evaluated at points of the curve. Indeed for every coordinate neighborhood on M we have frames F_i, $i = 1, \ldots, m$, and functions Γ_{ij}^k, which give $DF_i/\partial u^j$. From these data DY/dt can then be computed according to (2.9) by ordinary differentiation of the components of Y and coordinates of $p(t)$.

$\nabla_{X_p} Y$ and Differentiation of Vector Fields

We will now change our point of view slightly in deriving some consequences of formula (2.9). Let Y be a tangent vector field on M which is defined everywhere—not just along some curve. On the coordinate neighborhood U we write $Y = \sum_{k=1}^m b^k(u) F_k$. Let p be a point of U such that $\varphi(p) = (u_0^1, \ldots, u_0^m)$, and let X_p be a tangent vector at p, $X_p = \sum a^j F_{jp}$, a^j constant for $j = 1, \ldots, m$. Now choose any differentiable curve $p(t)$ whatsoever with $p(t_0) = p$ and $(dp/dt)_{t_0} = X_p$, so that in local coordinates it is defined by $u(t) = (u^1(t), \ldots, u^m(t))$ with $u^i(t_0) = u_0^i$ and $(du^i/dt)_{t_0} = a^i$. Then we may compute $(DY/dt)_{t=t_0}$ as above with a surprising result. First we observe that $Y(t) = \sum b^k(u(t)) F_k$ implies that

$$\left(\frac{db^k}{dt}\right)_{t_0} = \sum_{j=1}^{m} \left(\frac{\partial b^k}{\partial u^j}\right)_{u_0} a^j = X_p b^k.$$

Taking this into consideration, (2.9) may be written in modified form as

$$\textbf{(2.10)} \qquad \left(\frac{DY}{dt}\right)_{t_0} = \sum_k \left(X_p b^k + \sum_{i,\, j} \Gamma_{ij}^k(u_0) b^i(u_0) a^j \right) F_k .$$

A careful examination of this formula discloses the remarkable fact that the right-hand side does not depend on $p(t)$ but *only* on its tangent vector X_p at p. Since $(DY/dt)_{t_0}$ is a vector in $T_p(M)$, this formula defines a mapping of $T_p(M)$ to itself $X_p \to (DY/dt)_{t_0}$. We introduce the notation $\nabla_{X_p} Y$ for the image of X_p, that is, we define $\nabla_{X_p} Y = (DY/dt)_{t_0}$ along *any* curve $p(t)$ with $p(t_0) = p$ and $(dp/dt)_{t_0} = X_p$. We have defined previously a "directional derivative" $X_p f$ of a *function* f with respect to a vector X_p; what we have just now done is define in similar fashion a rate of change of the *vector field* Y at p in the direction X_p. It is measured by a vector $\nabla_{X_p} Y$.

It is worth commenting that along the curve $p(t)$ we have at each point $\nabla_{dp/dt} Y = DY/dt$ as a consequence of our notation. We summarize the essential properties of $\nabla_{X_p} Y$ in a theorem.

(2.11) Theorem *Let $M \subset \boldsymbol{R}^n$ be a submanifold. For any tangent vector field Y of class C^r, $r > 1$, on M we have at each point $p \in M$ a linear mapping $X_p \to \nabla_{X_p} Y$ of $T_p(M) \to T_p(M)$. Then $\nabla_{X_p} Y$, being defined as above, has the following properties:*

(1) *If X, Y are vector fields of class C^r (of class C^∞) on M, then $\nabla_X Y$ defined by $(\nabla_X Y)_p = \nabla_{X_p} Y$ is a C^{r-1} (respectively, C^∞) vector field on M.*

(2) *The map $T_p(M) \times \mathfrak{X}(M) \to T_p(M)$ given by $(X_p, Y) \to \nabla_{X_p} Y$ is $\boldsymbol{R}$-linear in X_p and Y. For a function f, differentiable on a neighborhood of p,*

$$\nabla_{X_p}(fY) = (X_p f)Y_p + f(p)\nabla_{X_p} Y.$$

(3) *If $X, Y \in \mathfrak{X}(M)$, then $[X, Y] = \nabla_X Y - \nabla_Y X$.*

(4) *If Y_1 and Y_2 are vector fields and (Y_1, Y_2) their inner product, then $X_p(Y_1, Y_2) = (\nabla_{X_p} Y_1, Y_{2p}) + (Y_{1p}, \nabla_{X_p} Y_2)$.*

Proof Let $Y = \sum b^k F_k$ and $X = \sum a^k F_k$ in the notation just used. The b^k are functions of the local coordinates $(u^1, \ldots, u^m)$ and so are the a^k when X is a vector field. Since $X_p b^k = \sum_{j=1}^m (\partial b^k/\partial u^j)a^j$, the definition $\nabla_{X_p} Y = DY/dt$ and (2.10) imply that

$$\nabla_{X_p} Y = \sum_k \sum_j \left(\frac{\partial b^k}{\partial u^j} a^j + \sum_i \Gamma_{ij}^k b^i a^j\right) F_k .$$

From this formula, valid for each $p \in U$, it is clear that properties (1) and (2) hold, whereas (4) is just the earlier property (2.3c) of DY/dt. [Again note that $X_p f = df/dt$, the derivative of $f(p(t))$, when we assume $X_p = dp/dt$; in particular this holds for $f = (Y_1, Y_2)$.] Only property (3) requires more careful verification. We will verify (3) by direct computation in a coordinate neighborhood U, φ using our previous notation. With X and Y given on U as above we compute $[X, Y]$:

$$[X, Y] = \sum_{k,j} \left(\frac{\partial b^k}{\partial u^j} a^j - \frac{\partial a^k}{\partial u^j} b^j\right) F_k .$$

Using the formula for $(DY/dt)_{t_0}$ we compute $\nabla_{X_p} Y - \nabla_{Y_p} X$. We have

$$\nabla_{X_p} Y - \nabla_{Y_p} X = \sum_k \left\{ \sum_j \left(\frac{\partial b^k}{\partial u^j} a^j - \sum \frac{\partial a^k}{\partial u^j} b^j \right) + \sum_{i,j} \Gamma^k_{ij}(b^i a^j - a^i b^j) \right\} F_k.$$

However, since $\Gamma^k_{ij} = \Gamma^k_{ji}$ this reduces to the first term in parentheses, the second sum being zero. Thus (3) follows. ∎

We conclude with several remarks. First, a careful reexamination of what we have done will show that $\nabla_{X_p} Y$ depends for its definition only on the Euclidean structure of $\boldsymbol{R}^n$, that is, on $\boldsymbol{E}^n$ and on the imbedding of M in $\boldsymbol{E}^n$. It is independent of local coordinates, although we use them in its definition and in the proof above. However, dY/dt and $DY/dt = \pi'(dY/dt)$ are geometric in nature and so is $\nabla_{X_p} Y$.

Secondly, if $\nabla_{X_p} Y$ is axiomatized and defined first, then DY/dt could be introduced by $DY/dt = \nabla_{dp/dt} Y$ and we could reverse our definitions and steps above.

Finally we note that although there is a partial duality of roles of X and Y in the symbol $\nabla_X Y$, which in fact defines an $\boldsymbol{R}$-bilinear mapping of $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ by $(X, Y) \to \nabla_X Y$, actually there is an important difference in the roles of X and Y. Namely this mapping is $C^\infty(M)$-linear in the first variable (Exercise 3) but not the second [by (2) of Theorem 2.11].

In this connection we observe that when X and Y are vector fields on M, then the Lie derivative $L_X Y = [X, Y]$ gives a rate of change or derivative of Y in the direction of X. However, this derivative requires a *vector field* X, not just a vector X_p at a single point as does $\nabla_{X_p} Y$. Thus the two concepts of differentiation are essentially different. Property (3) of Theorem 2.11 gives their precise relationship.

Exercises

1. Prove Lemma 2.1. [*Hint:* The ideas in the next exercise may help.]
2. Let U, φ be a coordinate neighborhood on $M \subset \boldsymbol{R}^n$ and let $F_1, \ldots, F_m$ be the coordinate frames on U—as in (2.6). If $p \in U$, then show that we may complete these frames to C^∞ frames $F_1, \ldots, F_{m+1}, \ldots, F_n$ of $\boldsymbol{R}^n$ on some neighborhood $V \subset U$ of p. (This means that the components of these vectors relative to the frames of $\boldsymbol{R}^n$ are C^∞ functions of the local coordinates of M.) Using the Gram–Schmidt orthogonalization process, show that there is a C^∞ orthonormal frame field $F'_1, \ldots, F'_n$ on this neighborhood V such that for each $k = 1, \ldots, n$, the vectors $F'_1, \ldots, F'_k$ and $F_1, \ldots, F_k$ span the same subspace. Use the new frames to give expressions for π', π'' on V.
3. Prove that if $X, Y \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, then $\nabla_{fX} Y = f \nabla_X Y$. Find the correct formula for $\nabla_X(fY)$.

4. Let M be a hypersurface of $\boldsymbol{R}^n$ (submanifold with dim $M = n - 1$). Show that if $p(t)$ is a curve on $M \subset \boldsymbol{R}^n$ and $Z(t) = Z_{p(t)}$ is a vector field along $p(t)$ such that $\|Z(t)\| \equiv 1$ and $Z(t)$ is always normal to M [that is, orthogonal to $T_{p(t)}(M)$], then DZ/dt is tangent to M.
5. Let M be the right circular cylinder $x^2 + y^2 = a^2$ in $\boldsymbol{R}^3$. Find all curves $p(t) = (x(t), y(t), z(t))$ on $\boldsymbol{R}^3$ which are geodesics, that is, for which $(D/dt)(dp/dt) = 0$.
6. Let $y = f(x)$, $a \leq x \leq b$, be a curve in the xy plane and let $M \subset \boldsymbol{R}^3$ be the surface obtained by revolving the curve around the x-axis. Assume $f(x)$ is C^2 at least and that $f(x) > 0$ on the interval. Show that the curve $y = f(x)$, $z = 0$, and each curve into which it rotates is a geodesic. Determine conditions for the intersection of M and a plane $x =$ constant to be a geodesic.

3 Differentiation on Riemannian Manifolds

We now pass to consideration of abstract Riemannian manifolds—manifolds which are not submanifolds of Euclidean space. Our purpose is to develop a satisfactory theory of differentiation on such manifolds, having properties like those discussed above but intrinsically defined, that is, without imbedding M in Euclidean space. We will reverse the order of ideas in the last section and *begin* by an attempt to define for M a derivative $\nabla_{X_p} Y$ of a vector field Y in the direction of a tangent vector X_p to M at p. Of course, we use the properties discovered in the previous section as our model. In all that follows we shall suppose all vector fields and functions on M to be C^∞.

(3.1) Definition A C^∞ *connection* ∇ on a manifold M is a mapping $\nabla: \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ denoted by $\nabla: (X, Y) \to \nabla_X Y$ which has the linearity properties: For all $f, g \in C^\infty(M)$ and $X, X', Y, Y' \in \mathfrak{X}(M)$, we have

(1) $\nabla_{fX+gX'} Y = f(\nabla_X Y) + g(\nabla_{X'} Y)$,
(2) $\nabla_X(fY + gY') = f\nabla_X Y + g\nabla_X Y' + (Xf)Y + (Xg)Y'$.

One should not attempt to read anything into the word "connection"—it is just an operator like the directional derivative. Note the asymmetry in the roles of the first and second vector fields X and Y; ∇ is $C^\infty(M)$ linear in X but not in Y. However, if f is a constant function, then $Xf = 0$; thus ∇ is $\boldsymbol{R}$-linear in both variables. Of course, we do not know that connections of this type exist, although by Theorem 2.11 they do for M imbedded in Euclidean space. In addition, according to that theorem we have in this special case two further properties:

(3) $[X, Y] = \nabla_X Y - \nabla_Y X$ (symmetry), and
(4) $X(Y, Y') = (\nabla_X Y, Y') + (Y, \nabla_X Y')$.

(3.2) Definition A C^∞ connection which also has properties (3) and (4) is called a *Riemannian connection.*

Note that in these definitions it is only property (4) that involves the Riemannian metric; thus on arbitrary differentiable manifolds one may study *C^∞ connections* [properties (1) and (2)] or *symmetric C^∞ connections* [properties (1)–(3)]. However, only the Riemannian case will be of interest to us.

(3.3) Theorem (Fundamental Theorem of Riemannian Geometry) *Let M be a Riemannian manifold. Then there exists a uniquely determined Riemannian connection on M.*

We prove this theorem in several steps in a manner somewhat similar to that of the existence proof for the operator d on $\bigwedge(M)$. Before doing so we deduce a consequence of the definition of connection which will resolve a minor discrepancy with the last section. In the discussion of differentiation on manifolds imbedded in $\boldsymbol{R}^n$ we defined the map $X_p \to \nabla_{X_p} Y$ from $T_p(M) \to T_p(M)$ using the vector field Y but without any assumption that X_p was the value at p of a vector field X. However, given vector fields X and Y, a vector field $\nabla_X Y$ was then defined by $(\nabla_X Y)_p = \nabla_{X_p} Y$ for $p \in M$, thus obtaining a map ∇ of pairs (X, Y) of vector fields to a vector field $\nabla_X Y$, as in our present definition. We have now taken this map on pairs of vector fields as the primary notion, and we wish to see that conversely, Y defines a linear map of $T_p(M) \to T_p(M)$ for each $p \in M$, that is, to see that $(\nabla_X Y)_p$ depends not on the vector *field* X but only on its *value* X_p *at* p. The same is not true of the dependence on Y as will appear in the corollary below.

(3.4) Lemma *Let $X, Y \in \mathfrak{X}(M)$ and suppose that either $X = 0$ or $Y = 0$ on an open set $U \subset M$. If ∇ is a connection [satisfying properties (1) and (2) of Definition 3.1], then the vector field $\nabla_X Y = 0$ on U.*

Proof Suppose that $Y = 0$ on U and $q \in U$. There is a C^∞ function f on M with $\operatorname{supp}(f) \subset U$ and $f \equiv 1$ on a neighborhood V of q (cf. proof of Theorem V.4.4). Note that $\bar{V} \subset \operatorname{supp}(f) \subset U$. Then since $Y = 0$ on U, $fY \equiv 0$ on M. However, property (2) implies that ∇_X takes the 0-vector field to 0; therefore $\nabla_X(fY) \equiv 0$ on M. But then, using property (2) again we have

$$0 = (\nabla_X(fY))_q = (X_q f)Y_q + f(q)(\nabla_X Y)_q = (\nabla_X Y)_q .$$

Since q is an arbitrary point of U, this completes the proof when $Y = 0$ on U. A parallel proof using property (1) applies when $X = 0$ on U. ∎

In fact, since $\nabla_X Y$ is $C^\infty(M)$-linear in X, essentially the same arguments imply a stronger property.

(3.5) Corollary *Let p be any point of M. If $X, X' \in \mathfrak{X}(M)$ such that $X_p = X'_p$, then for every vector field Y, $(\nabla_X Y)_p = (\nabla_{X'} Y)_p$. Denote this uniquely determined vector by $\nabla_{X_p} Y$. Then the mapping from $T_p(M) \to T_p(M)$ defined by $X_p \to \nabla_{X_p} Y$ is linear.*

Proof Let U, φ be a coordinate neighborhood of the point p. As in the proof of the lemma, there is a C^∞ function on M with $\operatorname{supp}(f) \subset U$ and $f \equiv 1$ on a neighborhood V of p (so $\bar{V} \subset U$). If $X \in \mathfrak{X}(M)$, then on U we have

$$X = \sum_{i=1}^{n} a_i E_i$$

with $a_i \in C^\infty(U)$ and $E_1, \ldots, E_n$ the vectors of the coordinate frames. We define $\tilde{X}, \tilde{E}_1, \ldots, \tilde{E}_n \in \mathfrak{X}(M)$ and $\tilde{a}_1, \ldots, \tilde{a}_m \in C^\infty(M)$ by $\tilde{X} = f^2 X$, $\tilde{E}_i = f E_i$ and $\tilde{a}_i = f a_i$, $i = 1, \ldots, n$, on U, and all to be zero (vectors and functions respectively) on the open set $M - \operatorname{supp}(f)$. Then we have

$$\tilde{X} = \tilde{a}_1 \tilde{E}_1 + \cdots + \tilde{a}_n \tilde{E}_n$$

on all of M; but on $\bar{V}$ this reduces to the equation above since $\tilde{X} = X$, $\tilde{E}_i = E_i$ and $\tilde{a}_i = a_i$ on this set. Applying Lemma 3.4 and property (1) of ∇ gives

$$\nabla_X Y = \nabla_{\tilde{X}} Y = \sum_{i=1}^{n} \tilde{a}_i \nabla_{\tilde{E}_i} Y \qquad \text{on} \quad V.$$

Hence

$$(\nabla_X Y)_p = \sum \tilde{a}_i(p)(\nabla_{\tilde{E}_i} Y)_p = \sum a_i(p)(\nabla_{E_i} Y)_p,$$

where the right-hand side depends *only* on the value X_p of the vector field X at p. This proves the first statement and the formula itself shows that the mapping $X_p \to \nabla_{X_p} Y = (\nabla_X Y)_p$ is a linear mapping of $T_p(M)$ into itself. For its value depends linearly on the components $a_1(p), \ldots, a_n(p)$ of X_p relative to the basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$. ∎

An important consequence of Lemma 3.4 is that it allows us to define (unambiguously) the *restriction* ∇^U of a connection ∇ defined on M to any open subset $U \subset M$. This is done as follows. Let X, Y be C^∞-vector fields on U and let $p \in U$. As in the proof of the corollary we define $\tilde{X}$, $\tilde{Y}$ to be C^∞-vector fields on M such that $\tilde{X} = X$ and $\tilde{Y} = Y$ on a neighborhood V of p. Then define $(\nabla_X^U Y)_q = (\nabla_{\tilde{X}} \tilde{Y})_p$, $q \in V$. A different choice of $\tilde{X}$, $\tilde{Y}$ gives a different V, but by the lemma the value of $\nabla_{\tilde{X}} \tilde{Y}$ will be the same on the intersection of the two V's. Since p is arbitrary, this defines a unique C^∞ vector field $\nabla_X Y$ on U. It is easily verified as an exercise that ∇^U is a connection and is Riemannian if ∇ is—using the induced Riemannian metric on U. The importance of this follows from the next lemma.

(3.6) Lemma *Suppose that a Riemannian connection ∇ exists for every Riemannian manifold. If it is unique for manifolds covered by a single coordinate neighborhood U, then it is unique for all manifolds. Conversely, if there exists a uniquely determined (Riemannian) connection ∇^U for every Riemannian manifold covered by a single coordinate neighborhood U, then there exists a uniquely determined Riemannian connection ∇ on every Riemannian manifold.*

Proof We suppose that ∇ is a Riemannian connection on M. By hypothesis there is a *uniquely* determined Riemannian connection $\tilde{\nabla}^U$ on each coordinate neighborhood U, φ of M (with the induced Riemannian metric). Let X, Y be vector fields on M and denote by X_U, Y_U their restrictions to U. It is an easy consequence of the definition of ∇^U, the restriction of ∇ to U, that $\nabla^U_{X_U} Y_U = (\nabla_X Y)_U$. Then on each coordinate neighborhood we have $(\nabla_X Y)_U = \tilde{\nabla}^U_{X_U} Y_U$ for, by the uniqueness assumption, $\tilde{\nabla}^U = \nabla^U$. Since M is covered by coordinate neighborhoods, this proves the first statement.

Now suppose that ∇^U is uniquely determined on every coordinate neighborhood U, φ of M. If there is defined on M a ∇ with properties (1)–(4) it must be unique by the above. We shall define ∇ on M as follows: Let $X, Y \in \mathfrak{X}(M)$ and let $p \in M$. Choose a coordinate neighborhood U, φ containing p and define $(\nabla_X Y)_U = \nabla^U_{X_U} Y_U$. This defines $\nabla_X Y$ not only at p but on the neighborhood U. It is easy to verify properties (1)–(4) since they hold for ∇^U. Suppose V, ψ is a coordinate neighborhood overlapping U; let $W = U \cap V$. Then W is a coordinate neighborhood using either coordinate map φ or ψ and, ∇^W being thus *uniquely* defined, we have at every point q of W

$$(\nabla^U_{X_U} Y_U)_q = (\nabla^W_{X_W} Y_W)_q = (\nabla^V_{X_V} Y_V)_q .$$

This completes the proof of the lemma. ∎

Proof of Theorem 3.3 The proof of the existence and uniqueness of a Riemannian symmetric connection, Theorem 3.3, is now reduced to the case of a manifold covered by a single coordinate neighborhood. Let U, φ cover the manifold M and let $x^1, \ldots, x^n$ denote the local coordinates and $E_1, \ldots, E_n$ the coordinate frames. Denoting the inner product by (X, Y) we have as components of the metric tensor the C^∞ functions $g_{ij}(q) = (E_{iq}, E_{jq})$ on $U = M$. The matrix $(g_{ij}(q))$ is symmetric, positive definite, hence it has a uniquely determined inverse $(g^{ij}(q))$ whose entries are C^∞ functions on U also. We shall show that there exists a unique Riemannian connection ∇ on M. First we note that if ∇ can be defined at all, then by properties (1) and (2) it is determined by the C^∞ functions Γ^k_{ij} on U, $1 \le i, j, k \le n$, defined by $\nabla_{E_i} E_j = \sum_{k=1}^n \Gamma^k_{ij} E_k$. In fact, if $X = \sum b^i(x) E_i$ and $Y = \sum a^j(x) E_j$ on U, then from (1) and (2) and our definition of Γ^k_{ij},

$$\nabla_X Y = \sum_k \left(X a^k + \sum_{i,j} \Gamma^k_{ij} a^j b^i \right) E_k .$$

Conversely, given functions Γ_{ij}^k on U, this formula defines a C^∞ connection satisfying (1) and (2).

However, the Γ_{ij}^k are not arbitrary C^∞ functions since a Riemannian connection satisfies the further properties (3) and (4). Because $[E_i, E_j] = 0$ for the coordinate frames, property (3) is equivalent to

$$0 = [E_i, E_j] = \nabla_{E_i} E_j - \nabla_{E_j} E_i = \sum_k (\Gamma_{ij}^k - \Gamma_{ji}^k) E_k$$

or, in fact, to the symmetry of Γ_{ij}^k in the lower indices:

(3′) $\Gamma_{ij}^k = \Gamma_{ji}^k$.

Finally, property (4) is equivalent to

$$E_k g_{ij} = E_k(E_i, E_j) = (\nabla_{E_k} E_i, E_j) + (E_i, \nabla_{E_k} E_j)$$

or to

(4′) $E_k g_{ij} = \sum_s (\Gamma_{ki}^s g_{sj} + \Gamma_{kj}^s g_{si}), \qquad 1 \le i, j, k \le n.$

Finally, using the matrix (g^{ij}) inverse to (g_{ij}), we define $\Gamma_{ijk} = \sum_s \Gamma_{ij}^s g_{sk}$, which implies that $\Gamma_{ij}^k = \sum \Gamma_{ijs} g^{sk}$. Thus the n^3 C^∞ functions Γ_{ij}^k determine the n^3 C^∞ functions Γ_{ijk} and conversely. Properties (3′) and (4′) become

(3″) $\Gamma_{ijk} = \Gamma_{jik}$

and

(4″) $\partial g_{ij}/\partial x^k = \Gamma_{kij} + \Gamma_{kji}$,

respectively, if we write $E_k g_{ij} = \partial g_{ij}/\partial x^k$, that is, if we consider g_{ij} as functions of the local coordinates.

In summary, given a Riemannian connection on M, covered by a single coordinate neighborhood, then if a Riemannian connection ∇ exists, it determines n^3 functions Γ_{ijk} of class C^∞ which have the two properties just mentioned. Conversely, it is easy to check by reversing these steps that any such functions determine a C^∞ Riemannian connection on M. Thus the theorem is completely established by the following lemma.

(3.7) Lemma *Let W be an open subset of $\mathbf{R}^n$ and let (g_{ij}) be a symmetric, positive definite matrix whose entries are C^∞ functions on W. Then there exists a unique family of C^∞ functions $\Gamma_{ijk}(x)$, $1 \le i, j, k \le n$ on W satisfying the two sets of equations* (3″) *and* (4″).

Proof Write Eq. (4″) twice more, each time permuting i, j, k cyclically. Then subtract the second of these equations from the sum of the first and third. Using (3″), $\Gamma_{ijk} = \Gamma_{jik}$, gives the unique solutions

$$\Gamma_{jki} = \frac{1}{2}\left(\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j}\right).$$

This completes the last step in the proof of the fundamental theorem 3.3. ∎

Suppose that U, φ is a local coordinate system with coordinates $x^1, \ldots, x^n$. Let $E_1, \ldots, E_n$ be the coordinate frames and $Y = \sum a^k E_k$ be the expression on U of the vector field Y. If $p \in U$ and $X_p = \sum b^k E_{kp}$, then we have the following formula for $\nabla_X Y$ on U.

(3.8) Corollary *For each $p \in U$, using the above notation, we have*

$$(\nabla_X Y)_p = \nabla_{X_p} Y = \sum_k \left(\sum_j b^j \frac{\partial a^k}{\partial x^j} + \sum_{i,j} \Gamma^k_{ij} a^i b^j\right) E_k$$

with

$$\Gamma^k_{ij} = \frac{1}{2} \sum_s g^{ks}\left(\frac{\partial g_{si}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^s} + \frac{\partial g_{js}}{\partial x^i}\right).$$

Proof As we have seen in the proof $(\nabla_X Y)_U$ is the same as $\nabla^U_{X_U} Y_U$, that is, ∇^U on X, Y restricted to U. For this reason we use the same symbol ∇ for all cases. The formula of the corollary follows at once from applying properties (1) and (2) defining a connection to $\nabla_{\sum b^j E_j}(\sum a^k E_k)$. ∎

Of course, this is the same formula we obtained earlier in the proof of Theorem 2.11 for a manifold M in Euclidean space. In fact we have an obvious corollary of the uniqueness of ∇:

(3.9) Corollary *In the case of an imbedded (or immersed) manifold in Euclidean space, the differentiation defined in Theorem 2.11 depends only on the Riemannian metric induced by the imbedding (but is otherwise independent of the imbedding).*

(3.10) Remark In Sections 1 and 2 we used the concept of differentiation of vector fields along curves dY/dt to define DY/dt and then $\nabla_X Y$ on submanifolds of $\boldsymbol{R}^n$. In this section we showed quite independently of the earlier discussion that there is a uniquely determined Riemannian connection ∇ on every Riemannian manifold M. Using this result we come full circle and

define—for a vector field Y and curve $p(t)$ on M—the covariant derivative DY/dt of $Y(t) = Y_{p(t)}$ by

$$\frac{DY}{dt} = \nabla_{dp/dt} Y.$$

Let Y be given locally by $Y = \sum b^k(x)E_k$ and $p(t)$ by $x(t) = (x^1(t), \ldots, x^n(t))$. Then from Corollary 3.8, with $X_{p(t)} = \sum \dot{x}^j(t)E_j = dp/dt$, it is easy to rederive formula (2.9) (renumbered here)

$$\textbf{(3.11)} \qquad \frac{DY}{dt} = \sum_{k=1}^{n} \left(\frac{db^k}{dt} + \sum_{i,j=1}^{n} \Gamma_{ij}^k(x(t))b^i(x(t)) \frac{dx^j}{dt} \right) E_k .$$

Since db^k/dt depend only on the values of $b^1, \ldots, b^n$, components of Y along the curve, the formula is valid when Y is defined only at points of the curve. Of course on any interval of the curve Y may be extended to a vector field on M, but DY/dt is independent of the extension by (3.11).

Constant Vector Fields and Parallel Displacement

A vector field Y on M is said to be *constant* if $\nabla_{X_p} Y = 0$ for all $p \in M$ and $X_p \in T_p(M)$. In general there do not exist such vector fields, even on small open subsets of M. However, given a differentiable curve $p(t)$, $0 \leq t \leq T$, there will be a vector field $X(t) = X_{p(t)}$ *constant* or *parallel* along $p(t)$ (by which we mean $DX/dt \equiv 0$).

(3.12) Theorem *Let $p = p(0)$, the initial point of the curve $p(t), 0 \leq t \leq T$, and let $X_p \in T_{p(0)}(M)$ be given arbitrarily. Then there exists a unique constant vector field $X_{p(t)}$ along $p(t)$ such that $X_{p(0)}$ has the given value. If $F_{1p}, \ldots, F_{np}$ is an orthonormal frame at $p(0)$, then there is a unique, parallel field of orthonormal frames on $p(t)$ which coincide with the given one at $p = p(0)$.*

Proof The proof depends on the existence theorem (Theorem IV.4.1). We shall first demonstrate that if t_0 is any point of the interval $[0, T]$, and $X_{p(t_0)}$ is an arbitrary tangent vector to M at $p(t_0)$, then there exists an ε-interval around t_0 such that there is a unique constant vector field $X_{p(t)}$ taking the value $X_{p(t_0)}$, at $p(t_0)$ and defined at each $p(t)$ on the intersection of $(t_0 - \varepsilon, t_0 + \varepsilon)$ and $[0, T]$. To see this, we first choose $\delta > 0$ such that for $t_0 - \delta < t < t_0 + \delta$, $p(t)$ lies in a single coordinate neighborhood V, φ. By (3.11), $X_{p(t)} = \sum_k a^k(t)E_{kp(t)}$ will have the desired properties on some interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ if and only if $a^k(t)$ satisfy

$$\frac{da^k}{dt} = -\sum_{i,j} \Gamma_{ij}^k \, a^i(t) \frac{dx^j}{dt}, \qquad k = 1, \ldots, n,$$

and the initial conditions $\sum a^k(t_0)E_{kp(t_0)} = X_{p(t_0)}$, as given. In this system of first order equations, the Γ_{ij}^k depend on t through $p(t)$, and $x^i(t) = \varphi(p(t))$ are

the local equations of the curve $p(t)$. By Theorem IV.4.1, functions $a^k(t)$ with these properties exist and are unique on some interval around t_0. Moreover, their differentiability class is at worst one less than that of $p(t)$ (since this is true of the derivatives dx^i/dt).

We further note that if an orthonormal frame $F_{1p(t_0)}, \ldots, F_{np(t_0)}$ is given at $p(t_0)$, it can be extended on $p(t)$ as n individual constant vector fields over some ε-interval around t_0. But then for each i, j

$$\frac{D}{dt}(F_i, F_j) = \left(\frac{DF_i}{dt}, F_j\right) + \left(F_i, \frac{DF_i}{dt}\right) = 0.$$

Hence $(F_{ip(t)}, F_{jp(t)})$ is constant along $p(t)$ on $t_0 - \varepsilon < t < t_0 + \varepsilon$. But since $(F_{ip(t_0)}, F_{jp(t_0)}) = \delta_{ij}$, this holds on the interval, i.e., $F_{1p(t)}, \ldots, F_{np(t)}$ is a constant orthonormal frame field on $(t_0 - \varepsilon, t_0 + \varepsilon)$.

It follows that beginning at $t_0 = 0$, the conclusions of the theorem follow for all $t < t_1$, for some $t_1 \le T$. If $t_1 < T$, this leads to a contradiction: We have, for example, $X_{p(t)}$, a constant vector field defined for all $0 \le t < t_1$. Let $F_{1p(t)}, \ldots, F_{np(t)}$ be a constant orthonormal frame field defined on $t_1 - \varepsilon < t < t_1 + \varepsilon$ for some $\varepsilon > 0$ and let $X_{p(t)} = \sum_k b^k(t)F_{kp(t)}$ for the overlapping set $(t_1 - \varepsilon, t_1)$. Since $X_{p(t)}$ and the frame field are both constant, the components $b^k(t)$ are constant; hence $X_{p(t)}$ may be extended to the entire interval $(t_1 - \varepsilon, t_1 + \varepsilon)$ by simply letting its components relative to the orthonormal frame field be these same constant values. Thus $t_1 = T$ and the result is proved. ∎

(3.13) Remark We remark that it is sufficient for the curve to be piecewise differentiable, for then we can move X_p along each piece separately. Therefore it follows from this theorem that given a piecewise differentiable curve $p(t)$, there exists an isomorphism, in fact isometry, $\tau_t \colon T_{p(0)}(M) \to T_{p(t)}(M)$ determined by the condition that $\tau_t(X_{p(0)})$ be a parallel (constant) vector field along $p(t)$. It is clear from our initial discussion of dX/dt along a curve $p(t)$ in Euclidean space that this would enable us to define the derivative of vector fields along curves on a Riemannian manifold M by comparing vectors at different points of the curve. The notion of parallel displacement along curves is sometimes taken as the starting point in studying differentiation on manifolds.

Exercises

1. Show that if $E_1, \ldots, E_n$ is a parallel frame field along a differentiable curve $p(t)$ in M and $X(t) = X_{p(t)}$ is a vector field along the curve defined by $X(t) = \sum_{i=1}^n a^i(t)E_{ip(t)}$, then

$$\frac{DX}{dt} = \sum_{i=1}^{n} \frac{da^i}{dt} E_{ip(t)}.$$

2. Using spherical coordinates we may cover the 2-sphere S of fixed radius a minus a single meridian from north to south pole by a single coordin-

ate system U, φ. The parameter mapping φ^{-1} takes $W = \{(u^1, u^2) \mid 0 < u^1 < 2\pi, 0 < u^2 < \pi\}$ onto the sphere as imbedded in $\boldsymbol{R}^3$ by φ:

$$\varphi^{-1}(u^1, u^2) = (a \cos u^1 \sin u^2, a \sin u^1 \sin u^2, a \cos u^2).$$

Find $g_{ij}(u^1, u^2)$ and compute $\Gamma^k_{ij}(u^1, u^2)$.

3. Let the upper half plane be considered as a manifold M covered by a single coordinate system $M = \{(x^1, x^2) \mid x^2 > 0\}$ with $U = M$ and coordinates (x^1, x^2). If $g_{ij}(x) = (x^2)^{-2}\,\delta_{ij}$, find Γ^k_{ij}.
4. Show that isometries of Riemannian manifolds preserve the Riemannian connection, that is, if $F\colon M_1 \to M_2$ is a diffeomorphism preserving the Riemannian metric, then $F_*(\nabla^{(1)}_X Y) = \nabla^{(2)}_{F_*(X)} F_*(Y)$.
5. Let M be a Riemannian manifold, W an open neighborhood of (u_0, v_0) on the uv plane, and $F\colon W \to M$ a C^∞ mapping. Let $\partial F/\partial u$ and $\partial F/\partial v$ denote the vectors tangent to the curves $v =$ constant and $u =$ constant, respectively, and let $D/\partial u$, $D/\partial v$ denote the covariant derivatives of any vector field along these respective curves. Using (3.11), show by direct computation that

$$\frac{D}{\partial v}\frac{\partial F}{\partial u} = \frac{D}{\partial u}\frac{\partial F}{\partial v}.$$

To which property of the Riemannian connection does this correspond?

4 Addenda to the Theory of Differentiation on a Manifold

In this section we insert a brief treatment of two topics which are closely related to the previous section, but which we do not need or use until the next chapter. First, we introduce the Riemann curvature tensor, and second, we briefly treat connections from the point of view of exterior differential forms.

The Curvature Tensor

It is a standard theorem of advanced calculus that second-order partial derivatives are independent of the order of differentiation:

$$\frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^j}\right) = \frac{\partial}{\partial x^j}\left(\frac{\partial f}{\partial x^i}\right).$$

For functions on manifolds the analogous property $X(Yf) = Y(Xf)$ does not hold in general. Indeed $[X, Y]$ measures the extent by which it fails:

$$X(Yf) - Y(Xf) = [X, Y]f.$$

[It still holds if $X = E_i$ and $Y = E_j$, since $E_k f$ may be identified with $\partial\hat{f}/\partial x^k$, $k = 1, \ldots, n$, if we allow $\hat{f}$ to denote the expression for the function on M in local coordinates $x^1, \ldots, x^n$.]

Since interchangeability of order of differentiation is measured by an interesting object $[X, Y]$ in the case of functions, it is natural to study the same question for ∇_X and ∇_Y derivatives of a vector field Z on M with respect to vector fields X, Y. A relatively simple example (Exercise 1) shows that in general $\nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) \neq 0$; hence it determines a vector field on M which may be thought of as analogous to $[X, Y]$. In fact, however, a more important expression, which involves also the measure of noninterchangeability of derivatives of functions $[X, Y]$, is the following related vector field, denoted by $R(X, Y)Z$ or $R(X, Y) \cdot Z$:

$$R(X, Y) \cdot Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X, Y]} Z. \tag{4.1}$$

It is readily verified that this formula defines a multilinear mapping of $\mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, that is, $R(X, Y) \cdot Z$ is $\boldsymbol{R}$-linear in each variable. However, from another point of view, in this expression $R(X, Y)$ is an *operator*, determined by the vector fields X and Y, and assigning to each vector field Z a new C^∞-vector field $R(X, Y) \cdot Z$. Note that if $[X, Y] = 0$, as is the case when $X = E_i$, $Y = E_j$ are vectors of a coordinate frame, then

$$R(X, Y) \cdot Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z),$$

so that if $R(X, Y) = 0$ on M, then ∇_{E_i} and ∇_{E_j} are interchangeable for all Z. A purely formal reason for the added term $-\nabla_{[X, Y]} Z$ in the definition is so that the following important theorem holds.

(4.2) Theorem *At any point p, the vector $(R(X, Y) \cdot Z)_p$ depends only on X_p, Y_p, Z_p, the values of the three vector fields at p, and not their values in a neighborhood or on M; thus formula (4.1) assigns to each pair of vectors $X_p, Y_p \in T_p(M)$ a linear transformation $R(X_p, Y_p)\colon T_p(M) \to T_p(M)$. In fact, $(X_p, Y_p) \to R(X_p, Y_p)$ is a linear mapping of $T_p(M) \times T_p(M)$ into the space of operators on $T_p(M)$.*

Proof From the definition of $R(X, Y) \cdot Z$ we see that it depends $\boldsymbol{R}$-linearly on each of the three arguments X, Y, Z. Moreover if f is a C^∞ function on M (not necessarily constant), we have

$$R(fX, Y) \cdot Z = R(X, fY) \cdot Z = R(X, Y) \cdot fZ = fR(X, Y) \cdot Z$$

as we may easily check by direct computation.

Now suppose that U, φ is a coordinate neighborhood. Let $(x^1, \ldots, x^n)$ denote the local coordinates and $E_1, \ldots, E_n$ the coordinate frames. We suppose that $X = \sum \alpha^i E_i$, $Y = \sum \beta^j E_j$, $Z = \sum \gamma^k E_k$. Then by the remarks above,

$$R(X, Y) \cdot Z = \sum_{i, j, k} \alpha^i \beta^j \gamma^k R(E_i, E_j) \cdot E_k$$

and we see that at a given point p of U the right-hand side involves first $R(E_i, E_j) \cdot E_k$, which is independent of the vector fields, and second the values of the functions α^i, β^i, γ^i only at the point p itself, not at nearby points. This proves the theorem. We used only properties (1) and (2) of the connection ∇, but the next fact uses the Riemannian metric. ∎

(4.3) Corollary *The formula* $R(X, Y, Z, W) = (R(X, Y) \cdot Z, W)$ *defines a* C^∞*-covariant tensor of order* 4. *This tensor depends only on the Riemannian metric on* M: *If* M_1, M_2 *are Riemannian manifolds and* $F: M_1 \to M_2$ *is an isometry, then* $F^*R_2 = R_1$.

Proof Since $R(X_p, Y_p) \cdot Z_p$ is defined as an element of $T_p(M)$ for any $p \in M$, its inner product $(R(X_p, Y_p) \cdot Z_p, W_p)$ with any $W_p \in T_p(M)$ is a well-defined real number. Thus for each p, $R_p(X_p, Y_p, Z_p, W_p) = (R(X_p, Y_p) \cdot Z_p, W_p)$ defines a multilinear function of four variables on $T_p(M)$, that is, an element of $\mathscr{T}^4(T_p(M))$. This clearly defines a C^∞-tensor field since both inner product and $R(X, Y) \cdot Z$ are C^∞ for $X, Y, Z, W \in \mathfrak{X}(M)$.

We have defined an *isometry* of Riemannian manifolds to be a diffeomorphism which preserves the Riemannian metric, that is, $F_*: T_p(M_1) \to T_{F(p)}(M_2)$ preserves inner products (and is an isomorphism onto). [If we do not suppose that the C^∞ mapping F is one-to-one onto, but only that F_* is onto and preserves inner products, then it is called a *local* isometry. It is an isometry on some neighborhood of each point (for example, covering spaces). The last statement of Corollary 4.3 is valid for local isometries also.] Now since ∇ is uniquely determined by the Riemannian metric, F_* preserves the connection, more precisely $F_*(\nabla^1_X Y) = \nabla^2_{F_*(X)} F_*(Y)$. From this we deduce that $R_2(F_* X, F_* Y) \cdot F_* Z = F_*(R_1(X, Y) \cdot Z)$. Since inner products are preserved, this implies $F^*R_2 = R_1$. ∎

(4.4) Definition The operator $R(X, Y)$ is called the *curvature operator* and the tensor $R(X, Y, Z, W)$ is called the *Riemann curvature tensor*. [It is not difficult to see that each one determines the other (Exercise 2).]

(4.5) Remark Let $E_1, \ldots, E_n$ be a field of frames on U, an open set of M. Then the Riemann curvature tensor is uniquely determined on U by either of the n^4 sets of functions R^j_{ikl} or R_{ijkl} defined by the equations

$$R(E_k, E_l)E_i = \sum_j R^j_{ikl} E_j$$

and

$$R(E_k, E_l, E_i, E_j) = R_{ijkl} = \sum_s g_{js} R^s_{ikl}, \qquad g_{js} = (E_j, E_s).$$

The Riemannian Connection and Exterior Differential Forms

There is another way of formulating the properties of covariant derivatives, connections, curvature tensor, and so on, which we shall now touch upon—it will be more fully treated later. We suppose that U is an open subset of a manifold M which has defined over it a field of C^∞ frames $E_1, \ldots, E_n$. The most usual case is when these are the coordinate frames of a coordinate neighborhood U, φ. However, in the case of a Riemannian manifold, which is our present interest, we might find it convenient to consider a neighborhood with orthonormal frames (Exercise 4). Corresponding to $E_1, \ldots, E_n$, we have at each $p \in U$ the dual basis $\theta^1, \ldots, \theta^n$ of $T_p^*(M)$, characterized by $\theta^i(E_j) = \delta^i_j$. It is a field of dual coframes on U and is clearly C^∞. If $\theta^1, \ldots, \theta^n$ are given, then conversely $E_1, \ldots, E_n$ are determined (Section V.1).

Now in defining $\nabla_X Y$ on a manifold so as to satisfy properties (1) and (2), we saw that it is enough to know $\nabla_{E_i} E_j$; for $\nabla_X Y$ may then be computed. In fact, $\nabla_{E_i} E_j = \sum_k \Gamma^k_{ij} E_k$, and we determined the Γ^k_{ij} above. If a connection is given so that Γ^k_{ij} are known on U, then we may define n^2 one-forms θ^k_j by $\theta^k_j = \sum_l \Gamma^k_{lj} \theta^l$. Conversely, given these one-forms, then $\Gamma^k_{ij} = \theta^k_j(E_i)$, and hence $\nabla_{E_i} E_j$, and the connection is determined. Indeed one checks at once that $\nabla_X E_j = \sum_k \theta^k_j(X) E_k$, that is, the values of the forms $\theta^1_j, \ldots, \theta^n_j$ on X are the components of $\nabla_X E_j$ relative to the given frames. Therefore, given U and $\theta^1, \ldots, \theta^n$ a field of coframes on U, then the connection is determined on U by the n^2 forms θ^k_j. They are called the *connection forms.*

Of course, the n^2 connection forms θ^k_j are not arbitrary, they must satisfy certain conditions corresponding to properties (1)–(4) of Definition 3.1 if they are to determine a connection on U, especially in the case of a Riemannian connection—the one we are interested in. We have the following restatement of the fundamental theorem of Riemannian geometry in terms of forms—although we restrict ourselves only to the case in which the manifold is covered by a single coordinate neighborhood or more precisely a neighborhood on which is defined a frame field.

(4.6) Theorem *Let M be a Riemannian manifold such that it has a covering by a C^∞ field of coframes $\theta^1, \ldots, \theta^n$. Then there exists a uniquely determined set of n^2 C^∞ one-forms θ^k_j, $1 \le j, k \le n$, on M satisfying the two equations*

(i) $d\theta^i - \sum_j \theta^j \wedge \theta^i_j = 0$,
(ii) $dg_{ij} = \sum_k (\theta^k_i g_{kj} + \theta^k_j g_{ki})$,

where $g_{ij} = (E_i, E_j)$, with $E_1, \ldots, E_n$ the uniquely determined field of frames dual to $\theta_1, \ldots, \theta_n$. The forms θ^k_j so determined define the Riemannian connection satisfying properties (1)–(4) *of the fundamental theorem by the formulas:*

(iii) $\nabla_X E_j = \sum \theta^k_j(X) E_k$, *and*
(iv) $\nabla_X(fY) = (Xf)Y + f\nabla_X Y, f \in C^\infty(M)$.

Conversely, the Riemannian connection determines θ_j^k as explained above and these θ_j^k satisfy (i) *and* (ii).

The proof is basically a duplication of that of the fundamental theorem [Theorem 3.3] in the case of manifolds covered by a single coordinate neighborhood (see Exercises 6–8). It can then be extended to general M in similar fashion.

If we recall that a Riemannian manifold M of the type described may be covered by an orthonormal frame field $E_1, \ldots, E_n$ with $g_{ij} = (E_i, E_j) = \delta_{ij}$, then we have a nicer version of the above. In this case we denote θ^i by ω^i and θ_j^i by ω_j^k. Using the fact that $g_{ij} \equiv \delta_{ij}$ (and hence $dg_{ij} = 0$), we obtain the following corollary:

(4.7) Corollary *Let M be a Riemannian manifold which has a covering by a field $\omega^1, \ldots, \omega^n$ of coframes whose dual frames $E_1, \ldots, E_n$ are orthonormal. Then there exists a unique set of n^2 one-forms ω_j^k, $1 \le j, k \le n$ on M satisfying*

(i) $d\omega^i - \sum_j \omega^j \wedge \omega_j^i = 0$,
(ii) $\omega_j^k + \omega_k^j = 0$.

These ω_j^k determine the Riemannian connection (as above) and conversely.

Finally, we note that since θ_j^k are uniquely determined by $\theta^1, \ldots, \theta^n$ and the Riemannian metric—the coframe field and the metric—then the exterior derivatives $d\theta_j^k$ are also uniquely determined, as are their expressions as linear combinations of the basis $\theta^i \wedge \theta^j$, $1 \le i < j \le n$, of two-forms on the domain U of $\theta^1, \ldots, \theta^n$. As we shall see in the next chapter, the coefficients in these linear combinations determine the components of the curvature tensor.

Exercises

1. Using $\nabla_{E_i} E_j = \sum_k \Gamma_{ij}^k E_k$ and Exercise 3.3, show that for the metric $g_{ij} = (x^2)^{-2}\delta_{ij}$ on $\{(x^1, x^2) | x^2 > 0\}$,
$$\nabla_{E_k}(\nabla_{E_i} E_j) - \nabla_{E_i}(\nabla_{E_k} E_j) \neq 0.$$
2. Show that $R(X, Y, Z, W)$ determines $R(X, Y) \cdot Z$, that is, if the values of the former on all vector fields are known, the same holds for the latter.
3. For a local coordinate system, compute R_{ikl}^j and R_{ijkl} in terms of Γ_{ij}^k and g_{ij}.
4. Show that $\omega^1, \ldots, \omega^n \in \bigwedge^1(M)$ satisfy the hypotheses of Corollary 4.7 if and only if the Riemannian metric is defined by the tensor $\sum_i \omega^i \otimes \omega^i$.
5. Suppose that $F\colon \mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is a $C^\infty(M)$-multilinear mapping of k vector fields $(X_1, \ldots, X_k)$ on M to a vector field

$F(X_1, \ldots, X_k)$ [$C^\infty(M)$ linear in each vector field separately], M a Riemannian manifold. Show that

$$\Phi(X_1, \ldots, X_k, X_{k+1}) = (F(X_1, \ldots, X_k), X_{k+1})$$

defines a covariant $(k+1)$ tensor.

6. Use Lemma V.8.4 to show that $[X, Y] = \nabla_X Y - \nabla_Y X$ is equivalent to property (i) of Theorem 4.6. Prove that property (ii) is equivalent to $X(Y, Z) = (\nabla_X Y, Z) + (Y, \nabla_X Z)$ if we use $Xf = df(X)$.
7. Prove directly, using differential forms, that there exists one and only one set of forms θ_j^k satisfying (i) and (ii) as asserted in Theorem 4.6.
8. Complete the proof of Theorem 4.6 using the results of Exercise 7 by showing that (iii) and (iv) define a Riemannian connection on M as claimed.

5 Geodesic Curves on Riemannian Manifolds

As a first example of the use of covariant differentiation on a Riemannian manifold—we shall define and study the class of curves called geodesics. Let $p(t)$ be a curve on M and dp/dt its velocity vector, defined for some open interval $a < t < b$ of $\mathbf{R}$; we suppose it to be of class C^2 at least.

(5.1) Definition The (parametrized) curve $p(t)$ is said to be a *geodesic* if its velocity vector is constant (parallel), that is, if it satisfies the condition $(D/dt)(dp/dt) = 0$, the *equation of a geodesic*, for $a < t < b$.

As we saw previously, when $M = \mathbf{R}^n$ with its usual metric this implies that the curve is a straight line. But in Section 2 it was seen that for a submanifold of $\mathbf{R}^n$ this can mean something quite different, an example being the great circles on $S^{n-1} \subset \mathbf{R}^n$.

The parameter on a geodesic is not arbitrary; the fact that a curve is a geodesic depends both on its shape and its parametrization as we may see from the example of a (geometric) straight line in $\mathbf{R}^2$ given parametrically by $x^1 = t^3$, $x^2 = t^3$. We write $p(t) = (t^3, t^3)$; then

$$\frac{dp}{dt} = 3t^2 \frac{\partial}{\partial x^1} + 3t^2 \frac{\partial}{\partial x^2}.$$

Since $D/dt = d/dt$ in $\mathbf{R}^2$, we have

$$\frac{D}{dt}\left(\frac{dp}{dt}\right) = \frac{D}{dt}\left(3t^2 \frac{D}{\partial x^1} + 3t^2 \frac{\partial}{\partial x^2}\right) = 6t \frac{\partial}{\partial x^1} + 6t \frac{\partial}{\partial x^2} \neq 0.$$

Therefore this curve is not a geodesic although the path traversed is the line $x^1 = x^2$. If $p(t)$, $0 < t < b$, is a nontrivial geodesic (not a single point), then

the permissible parametrizations—those with respect to which it remains a geodesic—are given by the following lemma.

(5.2) Lemma *Let $p(t)$, $a < t < b$, be a nontrivial geodesic and let t' be a new parameter. With respect to t' the curve will be a geodesic if and only if $t = ct' + d$, $c \neq 0$ and d constant. In particular, the arclength is always such a parameter.*

Proof If we introduce a new parameter t' by $t = ct' + d$, $c \neq 0$, then $dp/dt' = c\, dp/dt$ and $(D/dt')(dp/dt') = c^2(D/dt)(dp/dt) = 0$; so the curve remains a geodesic relative to t'. Now let s be arclength measured from $p(t_0)$, a point of the curve. Then $ds/dt = \|dp/dt\|$. Since dp/dt is constant along the curve, by (2.3c) its length $\|dp/dt\|$ is constant. Either $\|dp/dt\|$ is identically zero with $p(t)$ a single point and $s = 0$ or else $ds/dt = \|dp/dt\| = c$, a nonzero constant, and $s = ct + d$. This means that the curve is a geodesic when parametrized by arclength. Since any other permissible parameter is related to arclength by a similar (linear) relation, any two parameters are linearly related. ∎

In order to make general statements about geodesics on manifolds we shall need to study the defining equation in some detail using the existence theorem (IV.4.1). We can, however, give a few further examples by virtue of the following two observations. First, the equation of a geodesic imposes only a *local* condition on the curve. More precisely, if each point of a curve C has a neighborhood in which it may be written in the form $p(t)$, $a < t < b$, with $(D/dt)(dp/dt) = 0$, then it is a geodesic; for then, using arclength from some fixed point as parameter on all of C, it must satisfy the equation $(D/ds)(dp/ds) = 0$ over its entire length. Second, the property of being a geodesic is preserved by isometries because covariant differentiation is preserved and therefore so is parallelism of a vector field (for example, dp/dt) along a curve.

Now we let $\pi: \boldsymbol{R}^2 \to T^2$ be the standard covering discussed in Example III.6.15 and in Section III.9. We take $\boldsymbol{R}^2$ with its usual Riemannian metric. Since the covering transformations are translations, they are isometries of $\boldsymbol{R}^2$. It follows that we may define on T^2 a Riemannian metric which makes the projection π a local isometry, meaning that π_* is an isometry of each tangent space $T_p(\boldsymbol{R}^2)$ onto $T_{\pi(p)}(T^2)$. With this metric the geometry of T^2 is locally equivalent to that of Euclidean space. [This Riemannian metric should not be confused with the metric induced on a torus imbedded in $\boldsymbol{R}^3$ by the standard Riemannian metric of $\boldsymbol{R}^3$.] Combining our two observations, it follows that even a local isometry, as for example this map π, carries geodesics onto geodesics. This means that the images of straight lines of $\boldsymbol{R}^2$ on T^2 are geodesics of T^2 (Fig. VII.10). In particular, lines of rational slope map to closed geodesics on T^2, lines of irrational slope do not—they are

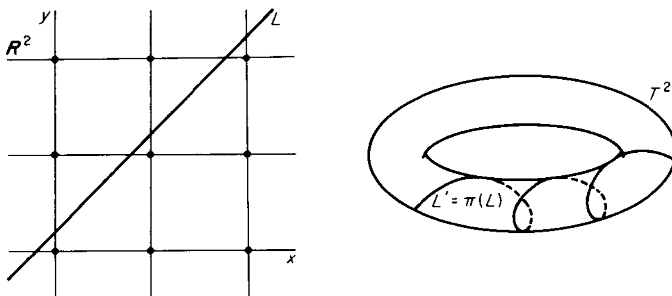


Figure VII.10

dense on T^2. Nothing like this occurs in $\boldsymbol{R}^2$, where geodesics can be neither closed curves nor dense. Thus "straight lines" even on spaces locally isometric to Euclidean space present some fascinating variations from what we might expect.

A similar, even simpler example is a right circular cylinder $M = \{(x, y, z) \in \boldsymbol{R}^3 \mid x^2 + y^2 = 1\}$. The Riemannian metric induced by that of $\boldsymbol{R}^3$ is the same as that given by the covering $\pi\colon \boldsymbol{R}^2 \to M$ (of Exercise 1). In this case the covering map is given by rolling up the plane into an infinite cylinder each strip of width 2π covering M once. Details are given in the exercises.

In order to study properties of geodesics on a Riemannian manifold M, we pass to local coordinates $(x^1, \dots, x^n)$ on a connected coordinate neighborhood U, φ. Then by (3.11) the equation of a geodesic $(D/dt)(dp/dt) = 0$ is equivalent to the system of second-order differential equations:

$$\text{(5.3)} \qquad \frac{d^2x^k}{dt^2} + \sum_{i,j=1}^{n} \Gamma^k_{ij}(x)\frac{dx^i}{dt}\frac{dx^j}{dt} = 0, \qquad k = 1, \dots, n.$$

A solution is a curve given in local coordinates by n functions $(x^1(t), \dots, x^n(t))$ which satisfy (5.3). As usual let $E_1, \dots, E_n$ denote the coordinate frames. We may apply our existence theorem IV.4.1 to prove the existence and uniqueness of a geodesic through each $p \in U$ with prescribed tangent direction at p and study its dependence on p and the tangent direction.

(5.4) Lemma *Given any $q \in U$, we can find a neighborhood V of q with $V \subset U$ and positive numbers r, δ such that for each $p \in V$ and each tangent vector $X_p = \sum b^i E_i$ with $\|X_p\| < r$, there exists a unique solution $(x^1(t), \dots, x^n(t))$ of (5.3), defined for $-\delta < t < \delta$, which satisfies $x^i(0) = x^i(p)$ and $\dot{x}^i(0) = b^i$, $i = 1, \dots, n$. Let $p(t) = \varphi^{-1}(x^1(t), \dots, x^n(t))$ as just defined. Then $p(t) \in U$ for $|t| < \delta$.*

Proof We consider the following system of $2n$ first-order ordinary differential equations defined on the open subset $W = \varphi(U) \times \boldsymbol{R}^n \subset \boldsymbol{R}^n \times \boldsymbol{R}^n = \boldsymbol{R}^{2n}$:

$$\text{(5.5)} \qquad \frac{dx^k}{dt} = y^k, \qquad k = 1, \ldots, n,$$

$$\frac{dy^k}{dt} = -\sum_{i,j=1}^{n} \Gamma_{ij}^k(x) y^i y^k, \qquad k = 1, \ldots, n.$$

The right-hand sides are C^∞ functions of $(x, y) = (x^1, \ldots, x^n; y^1, \ldots, y^n)$ on W. Therefore, according to the existence theorem for ordinary differential equations cited above, for each point in W there exists a $\delta > 0$ and a neighborhood $\tilde{V}$ of the point with the property that given $(a; b) = (a^1, \ldots, a^n; b^1, \ldots, b^n) \in \tilde{V}$, there are $2n$ unique functions $x^k = f^k(t, a; b)$ and $y^k = g^k(t, a; b)$, $k = 1, \ldots, n$ and $|t| < \delta$, satisfying the system of equations (5.5) and the initial conditions $f^k(0, a; b) = a^k$ and $g^k(0, a; b) = b^k$, $k = 1, \ldots, n$. These functions are C^∞ in all variables and have values in W. If $p \in U$, we consider the point $(\varphi(p); 0) = (x^1(p), \ldots, x^n(p); 0, \ldots, 0) \in W$. Then there is a $\delta > 0$ and a neighborhood $\tilde{V}$ of $(\varphi(p), 0)$ as described. This neighborhood may be chosen to be of the form $\varphi(V) \times B_{r'}^n(0)$ for some V with $\bar{V} \subset U$ compact and $r' > 0$. Since $\bar{V}$ is compact, we may find a number $r > 0$ such that if $(\sum g_{ij}(x) b_i b_j)^{1/2} = \|X_p\| < r$ and $p \in V$, then $(\sum (b^i)^2)^{1/2} < r'$. This follows from the inequalities used in the proof of Theorem V.3.1. We see at once from the special nature of system (5.5) that $df^k/dt = g^k$ and hence

$$\frac{d^2 f^k}{dt^2} = -\sum_{i,j} \Gamma_{ij}^k \frac{df^i}{dt} \frac{df^j}{dt}.$$

In other words $x^k(t) = f^k(t, a; b)$ are solutions of the system of equations (5.3), and therefore the equations in local coordinates of geodesics satisfying $x^k(0) = a^k$ and $(dx^k/dt)_{t=0} = b^k$, $k = 1, \ldots, n$. Finally, according to the existence theorem cited, the image of $I_\delta \times \tilde{V}$ under the map

$$(t, a, b) \to (f^1(t, a; b), \ldots, f^n(t, a; b); g^1(t, a; b), \ldots, g^n(t, a; b))$$

is in W which proves that $p(t) = \varphi^{-1}(f(t, a; b)) \in U$. ∎

The lemma has the following corollary, which guarantees the existence of a unique open geodesic arc through any given point with prescribed direction.

(5.6) Corollary *If M is a Riemannian manifold, $p \in M$, and Y_p is a nonzero tangent vector at p, then there is a $\lambda > 0$ and a geodesic curve $p(t)$ on M defined on some interval $-\delta < t < \delta$, $\delta > 0$, such that $p(0) = p$,*

$(dp/dt)_{t=0} = \lambda Y_p$. *Any two geodesic curves satisfying these two initial conditions coincide in a neighborhood of p.*

To see that this is so we take a neighborhood U, φ of p and choose $\lambda > 0$ so that $\|\lambda Y_p\| < r$ as in Lemma 5.4; then we apply the lemma.

(5.7) Remark Since "being a geodesic" is a purely local property of parametrized curves—meaning that they satisfy a differential equation—it is clear that if two geodesic curves C_1 and C_2 coincide (as sets) over some interval, then their union—suitably parametrized—is a geodesic. Further, we now see that if two geodesics have a single point in common and are tangent at that point, then their union is a geodesic. This implies that each geodesic is contained in a unique maximal geodesic. A *maximal geodesic* is one that is not a proper subset of any geodesic: If it is parametrized by a parameter t with $a < t < b$, then a and b (which can be $-\infty$ and/or $+\infty$) are determined by the curve and the choice of parameter. It is not possible to extend the definition of $p(t)$ (with the given parameter) so as to include either of these values and so that it will still be a geodesic. We shall be interested in determining conditions on M which ensure that $a = -\infty$ and $b = +\infty$ for every geodesic, or that every geodesic can be extended indefinitely in either direction. By Lemma 5.2 this property would be independent of parameter. It is easy to see that this is not always possible: let M be $\boldsymbol{R}^2$ with the origin removed. Then radial straight lines cannot be extended to the origin. However, given a geodesic through a point p, clearly we can always reparametrize it so that $p = p(0)$ and $p(t)$ is defined for $|t| < 2$, say. Making use of this fact, we modify Lemma 5.4 slightly to obtain our basic existence and uniqueness theorem for geodesics.

(5.8) Theorem *Let M be a Riemannian manifold and U, φ a coordinate neighborhood of M. If $q \in U$, then there exists a neighborhood V of q and an $\varepsilon > 0$ such that if $p \in V$ and $X_p \in T_p(M)$ with $\|X_p\| < \varepsilon$, then there is a unique geodesic $p(t) = p(t, p, X_p)$ defined for $-2 < t < +2$ and with $p(0) = p$, $(dp/dt)_{t=0} = X_p$. The mapping into M defined by $(t, p, X_p) \to p(t, p, X_p)$ is C^∞ on the open set $|t| < 2$, $p \in V$, $\|X_p\| < \varepsilon$ and has its values in U.*

Proof According to Lemma 5.4, we may find a neighborhood V of q and numbers $r, \delta > 0$ such that given any $p \in V$ and vector $X_p \in T_p(M)$ with $\|X_p\| < r$, then there is a geodesic $p(t)$ defined for $|t| < \delta$ and satisfying the initial conditions $p(0) = p$, $(dp/dt)_0 = X_p$. We know that if we change to a parameter $t = ct'$, $c \neq 0$ a constant, then $\tilde{p}(t') = p(ct')$ is again a geodesic with $\tilde{p}(0) = p$ and $d\tilde{p}/dt' = (dp/dt)(dt/dt') = c\, dp/dt$; thus $(d\tilde{p}/dt')_0 = cX_p$. Now if $\delta \geq 2$, we may use $\varepsilon = r$ and we have no more to prove; but if $\delta < 2$, we let $\varepsilon = \delta r/2$. Then, if $p \in V$ and X_p is a tangent vector at p with $\|X_p\| < \varepsilon$, we know from the choice of ε that $\|2X_p/\delta\| < r$. Thus there is a geodesic $p(t)$

with $p(0) = p$ and $(dp/dt)_0 = 2X_p/\delta$ defined for $|t| < \delta$ at least. The curve $\tilde{p}(t') = p(\delta t'/2)$ is again a geodesic and satisfies $\tilde{p}(0) = p$, $(d\tilde{p}/dt')_0 = (\delta/2)\times(dp/dt)_0 = X_p$. Moreover it is defined for $|\delta t'/2| < \delta$, that is, for $-2 < t' < +2$. This completes the proof; the last statement is already contained in Lemma 5.4. ∎

Exercises

1. Show that the mapping π: $(u, v) \to (\cos u, \sin u, v)$ carrying $\boldsymbol{R}^2$ onto the cylinder $M = \{(x, y, z) \mid x^2 + y^2 = 1\} \subset \boldsymbol{R}^3$ is a covering and a local isometry onto M with the induced metric of $\boldsymbol{R}^3$, from $\boldsymbol{R}^2$ with the usual metric.
2. Use Exercise 1 to show that the geodesics on M are the helices, that is, curves which cut each generator at the same angle (or have a constant angle with the z-axis), the generators themselves, and the circles of intersection with planes z = constant. Find how many geodesics connect two given points p, q.
3. Show that two isometries $F_1, F_2 \colon M \to M$ of a Riemannian manifold which agree on a point p and induce the same linear mapping on $T_p(M)$ agree on a neighborhood of p. Can you improve this statement?
4. Using the notation and Riemannian metric of Exercise 3.3, show that if suitably parametrized, the curves x^1 = constant are geodesics on M.

6 The Tangent Bundle and Exponential Mapping. Normal Coordinates

Although it may not have been apparent, the process by which we passed from a second-order system of equations (5.3) to a first-order system (5.5) when we first studied geodesics was to introduce new variables which corresponded to the components of tangent vectors at points of a coordinate neighborhood U, φ. These vectors X_p, $p \in U$, are in one-to-one correspondence with points $(x; y)$ of the open set $W = \varphi(U) \times \boldsymbol{R}^n \subset \boldsymbol{R}^n \times \boldsymbol{R}^n$. The correspondence, which we denote by $\tilde{\varphi}$, is given by $\tilde{\varphi}(X_p) = (\varphi(p); y^1, \ldots, y^n)$, where $\varphi(p) = (x^1, \ldots, x^n)$ are the coordinates of p, $X_p = \sum y^i E_{ip}$, and $E_1, \ldots, E_n$ are the coordinate frames. The differential equations of geodesics (5.3) were interpreted as a system of first-order differential equations (5.5) on W. Like all such systems, they correspond to a vector field on W (which we discuss in Section 7).

In order to free ourselves from working exclusively with local coordinates, it is natural to try to think of W being the image under $\tilde{\varphi}$ of a coordinate neighborhood $\tilde{U}, \tilde{\varphi}$ on a manifold. This is possible. It requires that we define a manifold structure on the set of *all* tangent vectors at *all* points of M, which we shall denote $T(M)$ (compare Section IV.2). When this is done,

$$T(M) = \{X_p \in T_p(M) \mid p \in M\} = \bigcup_{p \in M} T_p(M)$$

will become a space, in fact a C^∞ manifold, whose points are tangent vectors to M (compare Section I.5). In view of the introductory remarks, it is clear that we shall want the subset $\tilde{U}$ consisting of all X_p such that $p \in U$ to be a coordinate neighborhood with $\tilde{\varphi}$ as coordinate map and W as image, $\tilde{\varphi}\colon \tilde{U} \to W$. This virtually dictates the choice of topology and differentiable structure. Let $\pi\colon T(M) \to M$ be the natural mapping taking each vector to its initial point $\pi(X_p) = p$; then $\pi^{-1}(p) = T_p(M)$.

(6.1) Lemma *Let M be a C^∞-manifold of dimension n. There is a unique topology on $T(M)$ such that for each coordinate neighborhood U, φ of M, the set $\tilde{U} = \pi^{-1}(U)$ is an open set of $T(M)$ and $\tilde{\varphi}\colon \tilde{U} \to \varphi(U) \times R^n$, defined as above, is a homeomorphism. With this topology $T(M)$ is a topological manifold of dimension $2n$ and the neighborhoods $\tilde{U}$, $\tilde{\varphi}$ determine a C^∞-structure relative to which π is an (open) C^∞-mapping of $T(M)$ onto M.*

Proof Let U, φ and U', φ' be coordinate neighborhood on M such that $U \cap U' \neq \varnothing$; then $\tilde{U} \cap \tilde{U}' \neq \varnothing$. Comparing the coordinates of $p \in U \cap U'$ and the components of any $X_p \in T_p(M)$ relative to the two coordinate systems, we obtain the formulas for change of coordinates in $\tilde{U} \cap \tilde{U}'$:

$$\tilde{\varphi}' \circ \tilde{\varphi}^{-1}(x^1, \ldots, x^n; y^1, \ldots, y^n) = \left(f^1(x), \ldots, f^n(x); \sum_{i=1}^{n} y^i \frac{\partial f^1}{\partial x^i}, \ldots, \sum_{i=1}^{n} y^i \frac{\partial f^n}{\partial x^i}\right),$$

where $x'^i = f^i(x^1, \ldots, x^n)$, $i = 1, \ldots, n$, are the formulas for change of coordinates $\varphi' \circ \varphi^{-1}$ on $U \cap U'$ and the change of components is as in Corollary IV.1.8. These are easily seen to be diffeomorphisms of $\tilde{\varphi}(\tilde{U} \cap \tilde{U}')$ onto $\tilde{\varphi}'(\tilde{U} \cap \tilde{U}')$. The remainder of the verification is left as an exercise. Note that in local coordinates π corresponds to projection of $\boldsymbol{R}^n \times \boldsymbol{R}^n$ onto its first factor. We should also note that locally, on the domain $\tilde{U}$ of each coordinate neighborhood of the type above, $T(M)$ is a product manifold, that is, as an open submanifold of $T(M)$, $\tilde{U}$ is diffeomorphic to $\varphi(U) \times \boldsymbol{R}^n$. In the case of Euclidean space, U, φ may be taken to be all of $M = \boldsymbol{R}^n$ so that $T(\boldsymbol{R}^n)$ is diffeomorphic to $\boldsymbol{R}^n \times \boldsymbol{R}^n$. It is clear that for every manifold M, $\dim T(M) = 2 \dim M$. ∎

(6.2) Definition $T(M)$ with the topology and C^∞ structure just defined is called the *tangent bundle* of M, $\pi\colon T(M) \to M$ the natural projection.

Using Theorem 5.8, we may define Exp, the exponential mapping; its domain $\mathscr{D}$ is some subset of $T(M)$. It is a nontrivial matter to characterize exactly what this subset is. However, the range of Exp is M itself, thus $\operatorname{Exp}\colon \mathscr{D} \to M$ maps a *vector* X_p to a *point* of M. The name derives from the

exponential mapping of matrices (Section IV.6) for reasons which will be discussed later. Now let U, φ be a coordinate neighborhood of M and suppose $q \in U$. If a neighborhood V of q and an $\varepsilon > 0$ are chosen as in Theorem 5.8, then for each X_p with $p \in V$ and $\|X_p\| < \varepsilon$, or equivalently, in the open subset $\{X_p \mid p \in V, \|X_p\| < \varepsilon\}$ of $T(M)$, the geodesic $p(t)$ with $p(0) = p$ and $(dp/dt)_0 = X_p$ is defined for $|t| < 2$. On this open set of $T(M)$ we define Exp as follows.

(6.3) Definition Exp $X_p = p(1)$, that is, the image of X_p under the *exponential mapping* is defined to be that point on the unique geodesic determined by X_p at which the parameter takes the value $+1$.

Thus each $q \in M$ has a neighborhood V such that Exp is defined on the open subset $\{X_p | p \in V, \|X_p\| < \varepsilon\} \subset \pi^{-1}(V)$. (Note that ε depends on q and its neighborhood V.) This information on $\mathscr{D}$ may be restated as follows: Let M_0 be the submanifold of $T(M)$ consisting of all zero vectors $0_p, p \in M$. Then $p \to 0_p$ maps M onto M_0 diffeomorphically and $\pi: M_0 \to M$ is its inverse. The application of Theorem 5.8 then guarantees that the domain $\mathscr{D}$ of Exp contains an open neighborhood of M_0 in $T(M)$.

We also note that since $\|dp/dt\|$ is constant along a geodesic $p(t)$, its length L from $p(0)$ to $p(1)$ is

$$L = \int_0^1 \left\| \frac{dp}{dt} \right\| dt = \int_0^1 \|X_p\| \, dt = \|X_p\|.$$

Thus Exp X_p is the point on the unique geodesic $p(t)$ determined by X_p, whose distance from p along the geodesic is the length of X_p (see Fig. VII.11).

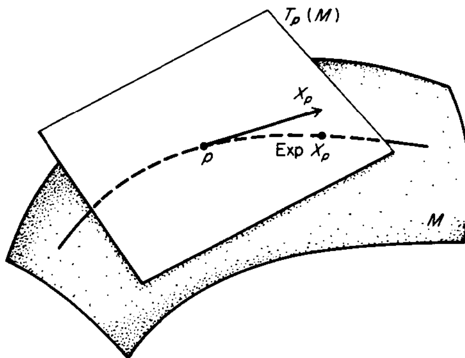


Figure VII.11

We shall use the following lemma to obtain identities (6.5):

(6.4) Lemma *Assume that $q \in M$ and that $X_q \in T_q(M)$ for which* Exp X_q *is defined. Then* Exp tX_q *is defined at least for each t with $0 \le t \le 1$ and $q(t) =$* Exp tX_q *is the geodesic through q at $t = 0$ with $(dq/dt)_0 = X_q$.*

Proof Let $q(t)$ be the unique geodesic with $q(0) = q$ and $(dq/dt)_0 = X_q$ so that Exp $X_q = q(1)$. Given c, $0 \le c \le 1$, consider the geodesic $\tilde{q}(t) = q(ct)$. We have $\tilde{q}(0) = q$ and $(d\tilde{q}/dt)_{t=0} = cX_q$ which means that Exp $cX_q = \tilde{q}(1) = q(c)$. Replacing c by t in this equality gives the statement above. ∎

We now revert once more to local coordinates U, φ around $q \in M$ and let $V \subset U$ and $\varepsilon > 0$ be as in Theorem 5.8 again so that for $p \in V$ and $\|X_p\| < \varepsilon$, Exp X_p is defined. As in the proof of Lemma 5.4, the geodesic determined by p, X_p is given in local coordinates by

$$t \to (f^1(t, a; b), \ldots, f^n(t, a; b))$$

with $\varphi(p) = a = (a^1, \ldots, a^n)$ and $X_p = b^1 E_{1p} + \cdots + b^n E_{np}$. This means that

$$\varphi(\text{Exp } X_p) = (f^1(1, a; b), \ldots, f^n(1, a; b))$$

and further that for $|t| < 1$

$$\varphi(\text{Exp } tX_p) = (f^1(1, a; tb), \ldots, f^n(1, a; tb)).$$

However, the Lemma 6.4 and the meaning of the functions $f^i(t, a; b)$ then give us the following identities valid for $|t| \le 1$:

$$\textbf{(6.5)} \qquad f^i(1, a^1, \ldots, a^n; tb^1, \ldots, tb^n) = f^i(t, a^1, \ldots, a^n; b^1, \ldots, b^n).$$

From these remarks we can draw some conclusions concerning the exponential map. First note that the f^i are C^∞ on their domain, hence $X_p \to \text{Exp } X_p$ is C^∞ on $\{X_p \mid p \in V, \|X_p\| < \varepsilon\}$. Second, we may compute the Jacobian of Exp_q at $X_q = 0_q$, the 0 vector at q, $q \in V$. We denote by Exp_q the restriction of Exp to $T_q(M) \cap \mathscr{D}$. Now q is fixed, $(a^1, \ldots, a^n)$ are constants, and the Jacobian matrix at this point has as entries $\partial f^i/\partial b^j$ evaluated at $(1, a^1, \ldots, a^n, 0, \ldots, 0)$:

$$\frac{\partial f^i}{\partial b^j} = \lim_{h \to 0} \frac{1}{h}(f^i(1, a; 0, \ldots, h, \ldots, 0) - f^i(1, a; 0, \ldots, 0))$$

Using the identities (6.5), with $b^j = 1$ and $b^k = 0$ for $k \neq j$, first with $t = h$, then with $t = 0$, this becomes

$$\begin{aligned}\frac{\partial f^i}{\partial b^j} &= \lim_{h \to 0} \frac{1}{h}(f^i(h, a; 0, \ldots, 1, \ldots, 0) - f^i(0, a; 0, \ldots, 1, \ldots, 0)) \\ &= \dot{f}^i(0, a^1, \ldots, a^n; 0, \ldots, 1, \ldots, 0).\end{aligned}$$

Since $x^i = f^i(t, a^1, \ldots, a^n; 0, \ldots, 1, \ldots, 0)$, $i = 1, \ldots, n$ (with $b^j = 1$ and $b^k = 0$ if $k \neq j$), considered as functions of t, are the equations of the geodesic through q with E_{jq} as initial vector, we see that the Jacobian matrix reduces to the identity at $X_q = 0_q$, that is; $\partial f^i/\partial b^j = \delta^i_j$. It follows that for q fixed and for some $\varepsilon' < \varepsilon$ the mapping $X_q \to \operatorname{Exp} X_q$ is a diffeomorphism of the open set $\tilde{N} = \{X_q \mid \|X_q\| < \varepsilon'\}$ of $T_q(M)$ onto an open set N containing $q = \operatorname{Exp} 0_q$. Retaining the notation Exp_q for Exp restricted to that part of its domain in $T_q(M)$, we summarize these results as follows.

(6.6) Normal Neighborhood Theorem *Every point q of a Riemannian manifold M has a neighborhood N which is the diffeomorphic image under* Exp_q *of a star-shaped neighborhood $\tilde{N}$ of the zero vector 0_q of the vector space $T_q(M)$.*

We have defined $\tilde{N}$ by $\|X_q\| < \varepsilon'$. Since the norm in $T_q(M)$ is given by the Riemannian metric, we may choose an orthonormal basis $F_1, \ldots, F_n$ of $T_q(M)$, and then, writing $X_q = \sum_{i=1}^n y^i F_i$, we have $\|X_q\|^2 = \sum_{i=1}^n (y^i)^2$. With these choices, the mapping

$$\psi: \operatorname{Exp}_q\left(\sum_{i=1}^n y^i F_i\right) \mapsto (y^1, \ldots, y^n)$$

takes the open neighborhood $N = \operatorname{Exp}_q(\tilde{N})$ of q diffeomorphically onto $B^n_{\varepsilon'}(0) \subset \mathbf{R}^n$.

(6.7) Definition The coordinate neighborhood N, ψ of q defined in this way is called a *normal coordinate* neighborhood.

(6.8) Remark Normal coordinates have special features that make them useful in the study of the geometry of the manifold. Of these the most important are the following:

(i) $g_{ij}(0) = \delta_{ij}$.

(ii) The equations of the geodesics through q take the form $y^i = a^i t$, $i = 1, \ldots, n$, a^i constants.

(iii) The coefficients of the connection vanish at q:

$$\Gamma^k_{ij}(0) = 0, \qquad i, j, k = 1, \ldots, n.$$

The first and second statements are immediate consequences of the definition and Lemma 6.4. The third follows from the second since for all $a^1, \ldots, a^n$ close to zero, substitution of the solutions $y^i = a^i t$ in the equations of the geodesics yields

$$\sum_{i,j} \Gamma^k_{ij}(0) a^i a^j = 0, \qquad k = 1, \ldots, n,$$

which implies (iii).

In fact the same computations used in the proof of the existence of normal neighborhoods give us a stronger result which we shall also find useful. We again let U, φ be a coordinate neighborhood of $q \in M$, let $E_1, \ldots, E_n$ denote the coordinate frames, $X_p = \sum b^i E_{ip}$ the tangent vectors to $p \in U$, and $\varphi(p) = (x^1, \ldots, x^n)$ the local coordinates. We have shown that there exists a relatively compact neighborhood V of q, $\bar{V} \subset U$, and an $\varepsilon > 0$ such that $\operatorname{Exp} X_p$ is defined and in U for each X_p with $p \in V$ and with $\|X_p\| < \varepsilon$. Then in local coordinates

$$\varphi(\operatorname{Exp} X_p) = (f^1(1, x^1, \ldots, x^n; b^1, \ldots, b^n), \ldots, f^n(1; x^1, \ldots, x^n, b^1, \ldots, b^n))$$

with $f^i(t, x, b)$ being C^∞ in all variables. We held p fixed at q to study the map Exp_q from $T_q(M)$ to M. Now, however, we consider the mapping F of the open set $\tilde{\varphi}(\{X_p | p \in V, \|X_p\| < \varepsilon\}) \subset \boldsymbol{R}^n \times \boldsymbol{R}^n$ to

$$\varphi(U) \times \varphi(U) \subset \boldsymbol{R}^n \times \boldsymbol{R}^n$$

which is defined by

$$F\colon (x^1, \ldots, x^n; b^1, \ldots, b^n) \mapsto (x^1, \ldots, x^n; f^1(1, x, b), \ldots, f^n(1, x, b)).$$

This map corresponds to the map $X_p = \sum b^i E_{ip} \to (p, \operatorname{Exp} X_p)$, with domain in $T(M)$. We have already seen that $\partial f^i / \partial b^j = \delta^i_j$ when $b^1 = \cdots = b^n = 0$. Therefore the Jacobian matrix of F is nonsingular at any point $(x^1, \ldots, x^n; 0, \ldots, 0)$ of $\boldsymbol{R}^n \times \{0\}$ for which $(x^1, \ldots, x^n) = \varphi(p)$ with $p \in V$. Thus by the inverse function theorem for each $0_p \in T(M)$, 0_p the zero vector at $p \in V$, there is a neighborhood which is mapped *diffeomorphically* onto an open subset of $U \times U \subset M \times M$ by this mapping, which takes the tangent vector X_p at p to a pair of points of U, $X_p \mapsto (p, \operatorname{Exp} X_p)$. Now V was originally chosen as a relatively compact neighborhood of q lying in a coordinate neighborhood U, φ. It was used to obtain an $\varepsilon > 0$ for which the open set $\{X_p | p \in V$ and $\|X_p\| < \varepsilon\}$ of $T(M)$ was in the domain $\mathscr{D}$ of Exp. This is also a set on which the mapping $X_p \to (p, \operatorname{Exp} X_p)$ is given in local coordinates by F. From what we have just said we may restrict V and ε further (without changing notation) so that the resulting neighborhood $\tilde{N}(V, \varepsilon) = \{X_p | p \in V$ and $\|X_p\| < \varepsilon\}$ of q, 0_q is mapped diffeomorphically onto an open set $W \subset U \times U$. Although W is not of the form $B \times B$, it does contain the diagonal set $\{(p, p) | p \in V\}$. We now let $B \subset V$ be a neighborhood of q such that $B \times B \subset W$. Then $B \times B$ is the diffeomorphic image of some open subset of $\tilde{N}(V, \varepsilon)$ which can be described by $\tilde{N}_B = \{X_p | p \in B, \operatorname{Exp} X_p \in B\}$. Putting these facts together gives the following result.

(6.9) Theorem *Let U, φ be a coordinate neighborhood of M and $q \in U$. Then there exists a neighborhood $B \subset U$ of q and an $\varepsilon > 0$ such that any two*

points p, p' of B can be joined by a unique geodesic of length less than ε. This geodesic is of the form $\operatorname{Exp} tX_p$, $0 \le t \le 1$, *and lies entirely in U. It follows that for each $p \in B$,* Exp_p *maps $\{X_p \mid \|X_p\| < \varepsilon\}$ diffeomorphically into an open set N_p such that $B \subset N_p \subset U$.*

We remark that our choice of the neighborhood $\tilde{N}_B$ does not allow us to conclude that whenever $X_p \in \tilde{N}_B$, then $\operatorname{Exp} tX_p \in B$ for all $0 < t < 1$. Thus in general B does not necessarily have the property that $p, p' \in B$ are joined by a geodesic lying entirely in B. We have made our choices so that for each $p \in V$, Exp_p maps the ε ball $\{X_p \mid \|X_p\| < \varepsilon\}$ into U diffeomorphically and clearly has B in its image, thus each $p \in B$ has a normal neighborhood N_p with $B \subset N_p \subset U$.

With somewhat more effort one can show that it is, in fact, possible to select a neighborhood B of each point q on a Riemannian manifold with the property that each pair of points $p, p' \in B$ may be joined by a unique (minimizing) geodesic segment lying entirely in B. Such neighborhoods are called *geodesically convex* and the proof of their existence is due to Whitehead [1]. It may be found in several of the references, for example, Helgason [1], Kobayashi and Nomizu [1], or Bishop and Crittenden [1].

Exercises

1. A *section* of $T(M)$ is a C^∞ mapping $F: M \to T(M)$ such that $\pi \circ F =$ identity on M. Prove that the sections of $T(M)$ correspond precisely to C^∞-vector fields on M.
2. Show that in a manner quite analogous to the definition of $T(M)$, a manifold structure can be defined on $\mathscr{T}^r(M)$ for any fixed r and that covariant tensor fields of order r correspond exactly to sections of M into $\mathscr{T}^r(M)$. (A "section" is defined as in Exercise 1.)
3. Show that the set of all unit tangent vectors at all points of M form a submanifold of $T(M)$. Discuss the existence of sections with image in this submanifold.
4. Let M be imbedded in $\boldsymbol{R}^m$ as a submanifold and for each $p \in M$ let $N_p \subset T_p(\boldsymbol{R}^m)$ be the subspace of vectors orthogonal to $T_p(M)$. Show that $N(M) = \bigcup_{p \in M} N_p$ can be given a structure of a C^∞ manifold such that the natural mapping $\pi: N(M) \to M$ given by mapping $N_p(M) \to p$ is C^∞. Proceed by analogy with $T(M)$.
5. Show that if G is a Lie group, then $T(G)$ is diffeomorphic to $G \times \boldsymbol{R}^n$, $n = \dim G$.
6. Let $F: M \to N$ be a C^∞ mapping of manifolds. Show that $F_*: T(M) \to T(N)$, defined by the usual mapping $F_*: T_p(M) \to T_p(N)$, is a C^∞ mapping of manifolds and commutes with the projection mappings.

7 Some Further Properties of Geodesics

Until now we have considered Exp only locally, restricted to a neighborhood of the zero vector at each point of the manifold. Our main purpose in this section is to prove, following Milnor [1], a theorem due to Hopf and Rinow [1] which gives conditions that the domain $\mathscr{D}$ of Exp be the entire tangent bundle $T(M)$. Equivalently, this means that Exp X_p is defined for every $p \in M$ and $X_p \in T_p(M)$. First, however, we wish to show that in all cases the domain $\mathscr{D}$ is an open set.

(7.1) Theorem *$\mathscr{D}$ is an open subset of $T(M)$ and* $\operatorname{Exp}: \mathscr{D} \to M$ *is a C^∞ mapping.*

Proof We adopt the notation of the previous section and recall that to each coordinate neighborhood U, φ of M corresponds a coordinate neighborhood $\tilde{U}, \tilde{\varphi}$ of $T(M)$. We have $\tilde{U} = \pi^{-1}(U)$ and $\tilde{\varphi}(\tilde{U}) = \varphi(U) \times \boldsymbol{R}^n \subset \boldsymbol{R}^n \times \boldsymbol{R}^n$. In fact, if $\varphi(p) = (x^1, \ldots, x^n)$ and $E_1, \ldots, E_n$ are the coordinate frames,

$$\tilde{\varphi}(X_p) = \tilde{\varphi}(\textstyle\sum y^i E_i) = (x^1, \ldots, x^n; y^1, \ldots, y^n).$$

The natural mapping $\pi: T(M) \to M$ is given in local coordinates by $\varphi(\pi(X_p)) = (x^1, \ldots, x^n)$; it is an open C^∞ mapping and has rank n at every point. Suppose that $p(t)$ is a geodesic on M. Then $X_{p(t)} = dp/dt$, its velocity vector, defines a curve $t \to X_{p(t)}$ on $T(M)$ with $\pi(X_{p(t)}) = p(t)$. An examination of the method by which we passed from the equations of geodesics (5.3) to first-order equations (5.5) reveals that on $\tilde{\varphi}(\tilde{U})$ (denoted by W in Lemma 5.4) we considered the first-order system corresponding to the vector field

$$Z' = \sum_i y^i \frac{\partial}{\partial x^i} - \sum_k \left(\sum_{i,j} \Gamma^k_{ij}(x) y^i y^j \right) \frac{\partial}{\partial y^k}.$$

Now we define a vector field Z on $\tilde{U} \subset T(M)$ so that $\tilde{\varphi}_*(Z) = Z'$. If, as in Lemma 5.4, the solutions of (5.5) are given by $x^i(t) = f^i(t, a, b)$ and $y^i(t) = dx^i/dt$, $i = 1, \ldots, n$, then on $\tilde{U}$ the integral curves (solutions) of the system of equations defined by Z are of the form

$$\tilde{\varphi}^{-1}\left(x^1(t), \ldots, x^n(t); \frac{dx^1}{dt}, \ldots, \frac{dx^n}{dt} \right),$$

where $\varphi^{-1}(x^1(t), \ldots, x^n(t)) = p(t)$ is a geodesic in $U = \pi(\tilde{U})$. In brief $X_{p(t)} = dp/dt$ is a solution curve of Z on $\pi^{-1}(U) \subset T(M)$ if and only if $p(t)$ is a geodesic on U. From its geometric meaning, or by a tedious computation for change of coordinates, we see that Z is a vector field defined intrinsically on all of $T(M)$, independent of the particular expression in a coordinate

system, that is, the components $(x^1, \ldots, x^n, -\sum_{i,j} \Gamma^1_{ij} y^i y^j, \ldots, -\sum_{i,j} \Gamma^n_{ij} y^i y^j)$ transform as they should for a vector field when we pass to other coordinates so that Z is globally defined and depends only on the Riemannian connection and metric. The geodesics on M are therefore exactly the projections by $\pi: T(M) \to M$ of the integral curves of Z. Thus the conclusion of the theorem follows from Theorem IV.4.5. [Note that IV (4.3)-(4.6) all hold!] ∎

We have seen that geodesics on Riemannian manifolds generalize straight lines in $\mathbf{R}^n$ in the following sense: Their unit tangent vector as we move along the curve is constant. But another basic property which characterizes straight lines in $\mathbf{R}^n$ is the famous minimizing property of being the shortest curve joining any two of its points. We now examine in some detail the extent to which this property generalizes. A few examples will show that there are some difficulties.

One of the more interesting is the right circular cylinder M with the Riemannian metric obtained by considering the plane $\mathbf{R}^2$ with its usual metric as universal covering (see Exercise 5.1). Then the geodesics on the cylinder are exactly those curves which go into straight lines if we roll the cylinder along the plane: vertical generators and helices. Thus two points not on a circle whose plane is orthogonal to the axis will be joined by an infinite number of distinct geodesics of different lengths (Fig. VII.12).

On S^2 the larger of the two arcs of a great circle which join two points p and q (which are not at opposite ends of at diameter) is not of minimal length, even among nearby circular arcs. Finally, for the plane with the

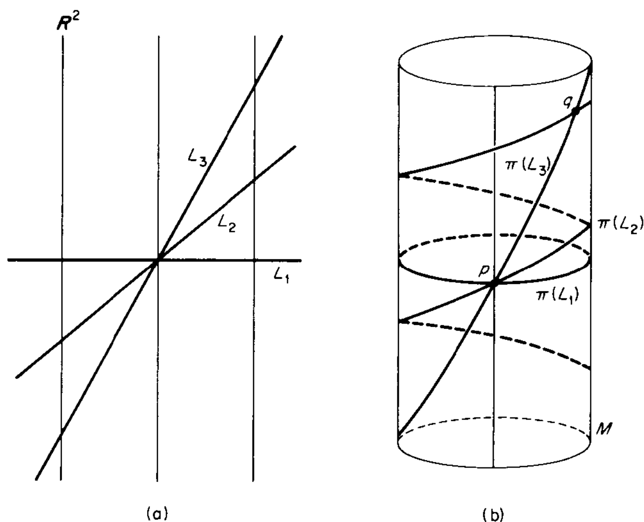


Figure VII.12

origin removed, the points $(-1, 0)$ and $(+1, 0)$ cannot be joined by a shortest curve at all. In view of all this it is remarkable that we are able to salvage something, in fact almost everything, if we limit ourselves to points close together and short geodesics. Let us recall that we have defined the length of a piecewise differentiable curve $p(t)$ (of class D^1) over $a \le t \le b$ by $L = \int_a^b \|dp/dt\| \, dt$; this is the Riemann integral of a piecewise continuous function. It is, by definition, equal to the sum of the integrals over the intervals of continuity [on each of which $p(t)$ is of class C^1]. Then we may elaborate Theorem 6.9 as follows.

(7.2) Theorem *For each $q \in M$, a Riemannian manifold, there exists a neighborhood B and an $\varepsilon > 0$ such that each pair of points of B can be joined by a unique geodesic of length $L < \varepsilon$. The length L' of any piecewise C^1 curve joining these two points is $\ge L$. Moreover $L' = L$ if and only if these paths coincide as point sets, or equivalently, when parametrized by arclength, are identical.*

This theorem is established using Theorem 6.9 and the following two lemmas. According to Theorem 6.9, given $q \in M$, there exists B and $\varepsilon > 0$ such that each pair of points p, p' of B can be joined by a unique geodesic of length $L < \varepsilon$. In fact, the equation $p(t)$ of the geodesic is given by $p(t) = \operatorname{Exp} tX_p$, $0 \le t \le 1$, and $\|X_p\| = L$. The open set B lies in a coordinate neighborhood U, φ which contains this geodesic, and Exp_p is a diffeomorphism of the open ball of vectors X_p of $T_p(M)$ of length $\|X_p\| < \varepsilon$ onto an open set N_p of U containing B. This means that any sphere $\{X_p \mid \|X_p\| = r < \varepsilon\}$ maps diffeomorphically to a submanifold of U, denoted by S_r (and called a *geodesic* sphere). The following lemma goes back to the work of Gauss.

(7.3) Lemma *Let $p \in B$ and suppose* Exp_p *maps the open ε-ball of $T_p(M)$ diffeomorphically onto $N_p \supset B$. Then the geodesics through p are orthogonal to the geodesic spheres S_r defined by $S_r = \{\operatorname{Exp}_p X_p \mid \|X_p\| = r\}$, $r < \varepsilon$.*

Proof Let $X(t)$ be a curve in $T_p(M)$ with $\|X(t)\| \equiv 1$, $a \le t \le b$. Any geodesic from the point p may be written $r \to \operatorname{Exp}_p rX$, $0 \le r \le \varepsilon$, with $\|X\| = 1$ and any curve on S_r in the form $t \to \operatorname{Exp}_p rX(t)$. The mapping $(r, t) \to p(r, t) = \operatorname{Exp}_p rX(t)$ maps the rectangle $[0, \varepsilon] \times [a, b]$ differentiably into M, and we will show that the inner product $(\partial p/\partial r, \partial p/\partial t) = 0$ for each r_0, t_0. These are the tangent vectors to $p(r, t_0)$, the geodesic curve, and to $p(r_0, t)$ a curve on the geodesic sphere S_r, respectively, intersecting at $p(r_0, t_0)$. If this inner product vanishes for every (r_0, t_0), this will establish

the lemma. We first show that $(\partial p/\partial r, \partial p/\partial t)$ is independent of r. By a basic property of differentiation,

$$\frac{D}{\partial r}\left(\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) = \left(\frac{D}{\partial r}\frac{\partial p}{\partial r}, \frac{\partial p}{\partial t}\right) + \left(\frac{\partial p}{\partial r}, \frac{D}{\partial r}\frac{\partial p}{\partial t}\right).$$

Of these the $(D/\partial r)(\partial p/\partial r) = 0$ since holding t fixed and allowing r to vary gives a geodesic through q with $\partial p/\partial r$ as its units tangent vecto. In the second term, if we interchange the order of differentiation (see Exercise 3.5), we obtain $(\partial p/\partial r, (D/\partial t)(\partial p/\partial r)) = \frac{1}{2}(D/\partial t)(\partial p/\partial r, \partial p/\partial r)$. Since $\|\partial p/\partial r\| = \|X(t)\| \equiv 1$, we see that this is also zero and therefore $(\partial p/\partial r, \partial p/\partial t)$ is independent of r. But $p(0, t) \equiv p$, so $\partial p/\partial t = 0$ at $r = 0$ and thus $(\partial p/\partial r, \partial p/\partial t) = 0$ for all r. Hence for each (r_0, t_0) the inner product $(\partial p/\partial r, \partial p/\partial t) = 0$, which completes the proof.

Now we consider a (piecewise) differentiable curve $\tilde{p}(t)$, $a \le t \le b$, in $N_p - \{p\}$; it has a unique expression of the form $\tilde{p}(t) = \operatorname{Exp}_p r(t)X(t)$, where $\|X(t)\| \equiv 1$. Using this notation, we state the following lemma.

(7.4) Lemma $\int_a^b \|d\tilde{p}/dt\|\, dt \ge |r(b) - r(a)|$. *Equality holds if and only if $r(t)$ is monotone and $X(t)$ is constant.*

Proof Again we consider the map $(r, t) \to p(r, t) = \operatorname{Exp}_p rX(t)$ from $[0, \varepsilon] \times [a, b] \to U$. The curve $\tilde{p}(t)$ connects the spherical shells S_r of radius $r = r(a)$ and $r = r(b)$ in U_q. We have $\tilde{p}(t) = p(r(t), t)$ and, using r' for dr/dt,

$$\frac{d\tilde{p}}{dt} = \frac{\partial p}{\partial r} r'(t) + \frac{\partial p}{\partial t}.$$

Since $\|\partial p/\partial r = \|X(t)\| = 1$ and $(\partial p/\partial r, \partial p/\partial t) = 0$ (Lemma 7.3), we have $\|d\tilde{p}/dt\|^2 = |r'(t)|^2 + \|\partial p/\partial t\|^2 \ge |r'(t)|^2$. Equality holds if and only if $\partial p/\partial t \equiv 0$, that is, $X(t) = $ constant. Hence,

$$\int_a^b \left\|\frac{d\tilde{p}}{dt}\right\| dt \ge \int_a^b |r'(t)|\, dt \ge \left|\int_a^b r'(t)\, dt\right| = |r(b) - r(a)|.$$

In the last inequality, we have equality only if $r(t)$ is monotone; thus $\int_a^b \|d\tilde{p}/dt\|\, dt = |r(b) - r(a)|$ if and only if $r(t)$ is monotone and $X(t) =$ constant. This proves the lemma.

Proof of Theorem 7.2 We continue the notation of the lemmas. Suppose $\tilde{p}(t)$, $0 \le t \le 1$, is a piecewise smooth curve joining $p = \tilde{p}(0)$ to $p' = \tilde{p}(1) = \operatorname{Exp}_p rX_p \in N_p$, $0 < r < \varepsilon$ and $\|X_p\| = 1$. Let δ satisfy $0 < \delta < r$, and consider the segment of the curve joining the shell of radius δ around p to that of radius r. According to Lemma 7.4, the length of this segment is $\ge r - \delta$ with equality holding only if the curve coincides as a point set with segment of the

radial geodesic from p cut off by these shells, its length being $r - \delta$. Thus the portion of the curve between these shells has strictly greater length than $r - \delta$ unless it coincides as a point set with a radial geodesic. Letting δ approach zero gives the result of the theorem. ∎

It is probably quite obvious that the statement of Theorem 7.2 is bound up with the notion of distance on M, that is, the metric $d(p, p')$ which we considered in Section V.3. Recall that $d(p, p')$ is the infimum of the lengths of all piecewise differentiable curves from p to p' and that we showed that the metric topology and the usual topology coincided. The theorem just proved guarantees that for each point $q \in M$ there is an $\varepsilon > 0$ and a neighborhood B of diameter less than ε (in terms of d) such that for every pair of points $p, p' \in B$ there is a unique geodesic segment from p to p' whose length is the distance $d(p, p')$. More generally, we have the corollary which follows.

(7.5) Corollary *If a piecewise differentiable path (of class D^1) from p to q on M has length equal to $d(p, q)$, then it is a geodesic when parametrized by arclength.*

Note that it follows that the path is C^∞! Of course the hypothesis and the definition of $d(p, q)$ imply that the path has minimum length among all such curves. The proof is immediate: any segment of the path lying in a sufficiently small neighborhood (as above) must also have as length the distance between its endpoints (or it could be replaced by a shorter path) and thus it must be a geodesic. Since the curve is a geodesic locally, it is a geodesic.

(7.6) Definition A geodesic segment whose length is the distance between its endpoints is called a *minimal* (or *minimizing*) *geodesic*.

Unlike the local situation in Theorem 7.2, we have seen that on an arbitrary manifold there may be points p, q which are not connected by a geodesic at all, for example, $\boldsymbol{R}^2$ with the origin removed. Moreover, even if there exist such minimal geodesics joining p, q as there do on the sphere, they need not be unique—for example, there are an infinite number of minimal geodesics joining the north and south poles. The question of uniqueness is not simple and we will not go into it here. For details, as well as many additional theorems on geodesics, the reader should consult Milnor [1].

However, the existence question as well as some other questions we have raised are answered in a beautiful theorem of Hopf and Rinow [1]. Before stating this theorem we remember that each geodesic and geodesic segment is contained in a maximal geodesic, that is, a geodesic $p(t)$ such that $p(t)$ is

defined for $a < t < b$ and not for any larger interval of values. If $a = -\infty$ and $b = +\infty$, we say that the geodesic can be *extended indefinitely*. This is always true of a *closed* geodesic (a geodesic which is the image of a circle, for example, a great circle on S^2). If every geodesic from $p \in M$ can be extended indefinitely, then the domain $\mathscr{D}$ of Exp contains all of $T_p(M)$ and conversely.

(7.7) Theorem (Hopf and Rinow) *Let M be a connected Riemannian manifold. Then the following two properties are equivalent:*

(i) *Any geodesic segment can be extended indefinitely.*
(ii) *With the metric $d(p, q)$, M is a complete metric space.*

The proof will be based on a lemma. Assume any geodesic segment $t \to p(t)$, $a \leq t \leq b$, can be extended to a maximal geodesic curve $t \to p(t)$, defined for $-\infty < t < +\infty$. In order to see that M is complete (every Cauchy sequence converges), it is sufficient to show that every closed and bounded set is compact; and to prove this we need the following lemma, which is of interest in itself. The proof is modeled on that of Milnor [1].

(7.8) Lemma *If M has the property that every geodesic from some point $p \in M$ can be extended indefinitely, then any point q of M can be joined to p by a minimal geodesic [whose length is necessarily $d(p, q)$].*

Proof Let q be an arbitrary point of M and let $a = d(p, q)$. Any geodesic from p may be written $p(s) = \operatorname{Exp} sX_p$ with X_p a unit tangent vector at p and s arclength measured from $p = p(0)$. We must show that for some X_p with $\|X_p\| = 1$, $p(a) = \operatorname{Exp} aX_p = q$, so that $s \mapsto \operatorname{Exp} sX$, $0 \leq s \leq a$, is the minimal geodesic segment. We will use the following fact, which is also of some interest.

(7.9) *Suppose that $p_0, p_1, \ldots, p_n$ are points of M and that*

$$(*) \qquad d(p_0, p_1) + d(p_1, p_2) + \cdots + d(p_{n-1}, p_n) = d(p_0, p_n).$$

If a piecewise differentiable curve contains $p_i, p_{i+1}, \ldots, p_{i+r}$ and has length equal to $d(p_i, p_{i+1}) + \cdots + d(p_{i+r-1}, p_{i+r})$, then it is a geodesic segment from p_i to p_{i+r}. Conversely, if *$p_0, \ldots, p_n$ lie on a minimal geodesic segment, in that order, then $(*)$ holds for them.*

It is easily seen that it is enough to verify this for $r = 2$. The curve C from p_i to p_{i+1} to p_{i+2} has length $L = d(p_i, p_{i+1}) + d(p_{i+1}, p_{i+2})$. By the triangle inequality $L \geq d(p_i, p_{i+2})$. If equality holds, C is a (minimizing) geodesic segment from p_i to p_{i+2} as required (Corollary 7.5). But this must be the case; otherwise we have

$$d(p_i, p_{i+1}) + d(p_{i+1}, p_{i+2}) > d(p_i, p_{i+2})$$

by the triangle inequality and then substituting in (*) we have

$$d(p_0, p_1) + \cdots + d(p_i, p_{i+2}) + \cdots + d(p_{n-1}, p_n) < d(p_0, p_n),$$

which contradicts the triangle inequality. Finally, the last statement follows immediately from the fact that any subsegment of a minimal geodesic segment is also minimal.

To return to the proof of Lemma 7.8, using Theorem 6.9 we suppose $\delta > 0$ to be chosen so that $S_\delta = \{p' \mid d(p, p') = \delta\}$ is a geodesic sphere in some normal neighborhood of p, sufficiently small to ensure that each radial geodesic from p to S_δ is minimal. Then since S_δ is compact, there is a $p_0 \in S_\delta$ satisfying

$$d(p_0, q) = \inf_{p' \in S_\delta} d(p', q).$$

Let X_p be the unit vector at p such that $p_0 = \operatorname{Exp} \delta X_p$. We must have

$$d(p, p_0) + d(p_0, q) = d(p, q),$$

otherwise there is a piecewise differentiable curve joining p to q whose length is less than $d(p, p_0) + d(p_0, q) = \delta + d(p_0, q)$. Since it must intersect S_δ at some point p' and its length from p to p' can be no less than δ, we have $d(p', q) < d(p_0, q)$ contrary to our choice of p_0. We now consider all s', $0 \leq s' \leq a$, such that the geodesic segment $s \mapsto \operatorname{Exp} sX_p$, $0 \leq s \leq s'$, is minimizing and such that

$$d(p, \operatorname{Exp} s'X_p) + d(\operatorname{Exp} s'X_p, q) = d(p, q).$$

By the continuity of the conditions the collection of all such s' forms a closed interval $0 \leq s' \leq b$. If $b = a$, then $\operatorname{Exp} aX_p = q$, which proves the lemma. Suppose $b < a$; let $p_1 = \operatorname{Exp} bX_p$, then $d(p, p_1) + d(p_1, q) = d(p, q)$ and we may obtain a contradiction by repeating the arguments above as follows. Let S_η, $\eta > 0$, be a small geodesic sphere (with radial geodesics minimizing) in a normal neighborhood of $p_1 = \operatorname{Exp} bX_p$ and choose a point p_2 on S_η such that

$$d(p_2, q) = \inf_{p'' \in S_\eta} d(p'', q).$$

Then, as before $d(p_1, p_2) + d(p_2, q) = d(p_1, q)$ and therefore

$$d(p, p_1) + d(p_1, p_2) + d(p_2, q) = d(p, p_1) + d(p_1, q) = d(p, q).$$

By (7.9) the geodesic $p(s) = \operatorname{Exp} sX_p$ from p to p_1 together with the (radial) geodesic in S_η from p_1 to p_2 is a single (minimizing) geodesic segment from p to p_2 of length $d(p, p_2) > b$, which contradicts the definition of b. Therefore $b = a$ and the lemma follows. ∎

Proof of Theorem 7.7 We now turn to the proof of the theorem. We show first that (i) implies (ii). Let K be a closed and bounded subset of M. We will show that K is compact. Suppose $p \in K$ and $a = \sup_{q \in K} d(p, q)$; a is finite since K is bounded. By Lemma 7.8, for any $q \in K$ there is a minimizing geodesic from p to q; its length is $d(p, q)$ which must be no greater than a. It follows that $K \subset \operatorname{Exp}_p \bar{B}_a$, where $\bar{B}_a = \{Y_p \mid \|Y_p\| \le a\}$, the closed ball of radius a in T_p. Since $\bar{B}_a$ is compact and Exp is continuous, $\operatorname{Exp}_p \bar{B}_a$ is compact. K is a closed subset of $\operatorname{Exp}_p \bar{B}_a$, so it must be compact. This completes the proof that M is a complete metric space according to Exercise 6.

(7.10) Remark Any manifold M having property (i) has the property that the domain $\mathscr{D}$ of the exponential function is all of $T(M)$, that is, that the vector field Z of Theorem 7.1 is complete. Actually, in proving that (i) implies (ii), we used only the weaker hypothesis of the lemma: every geodesic from *some* point $p \in M$ can be extended indefinitely, that is, $\mathscr{D} \supset T_p(M)$ for some $p \in M$. It was not necessary to assume $p \in K$, for if K is bounded, then for any $p \in M$ the distances $d(p, q)$ are bounded for all $q \in K$.

Next we show that (ii) implies (i), that is, we suppose that every Cauchy sequence on M converges and show that this implies the extendability of geodesics. Suppose to the contrary that there is a geodesic ray, $p(t)$, $0 \le t < t_0$ which cannot be extended to $t = t_0$; we may assume, changing parameter if necessary, that t is arclength. Let $\{t_n\}$ be an increasing sequence of parameter values with $\lim_{n \to \infty} t_n = t_0$. Denoting by p_n the points $p(t_n)$, we have $d(p_n, p_m) \le |t_n - t_m|$ since the right-hand side is the length of a curve (the geodesic) from p_n to p_m. Thus $\{p_n\}$ is a Cauchy sequence and we denote its limit by q, $q = \lim_{n \to \infty} p_n = \lim_{t_n \to t_0} p(t_n)$. Now we let B be a neighborhood of q, and $\varepsilon > 0$ be so chosen that each pair of points p, p' of B are joined by a unique geodesic of length less than ε. Of course this geodesic is minimizing, or equivalently its length is $d(p, p')$. We let N be an integer which is large enough so that for $n, m \ge N$ we have $d(p_n, p_m) < \varepsilon$ and $d(p_n, q) < \varepsilon$ and $p_n, p_m \in B$.

Consider $n \ge N$ fixed and suppose $m > n$; then we have

$$d(p_n, p_m) + d(p_m, q) = (t_m - t_n) + d(p_m, q).$$

Since $t_m - t_n$ is the length of our geodesic from p_n to p_m and is less than ε, this segment of the geodesic is minimal. Now let $m \to \infty$ and by continuity we have $d(p_n, q) = t_0 - t_n$ for $n > N$. Applying this to $m > n$, we have for all $m > n > N$,

$$d(p_n, p_m) + d(p_m, q) = t_m - t_n + t_0 - t_m = t_0 - t_n = d(p_n, q).$$

Now choosing a fixed $m > n$, we see that the unique geodesic segment from p_n to p_m of length $d(p_n, p_m)$ together with the unique geodesic segment from p_n to q of length $d(p_n, q)$ has length equal to the distance $d(p_n, q)$ and therefore is a single (unbroken) geodesic from p_n to q. However, it coincides with the given geodesic $p(t)$ for $t_n \le t \le t_m$, that is, from p_n to p_m; thus it is an extension of this to a geodesic segment from p to q. This shows that $p(t)$ can be extended to $t = t_0$.

We note that it is immediate that a geodesic segment $p(t)$, $0 \le t \le t_0$, can be extended beyond its endpoints; this follows at once from the fundamental existence theorems. Thus any geodesic on a complete manifold can be extended indefinitely, Exp_p is defined on all of $T_p(M)$ for every p, and Exp has the entire tangent bundle $T(M)$ as its domain, that is, $\mathscr{D} = T(M)$. ∎

The following corollary depends on the fact that a compact metric space is complete.

(7.11) Corollary *If a connected Riemannian manifold M is compact, then any pair of points $p, q \in M$ may be joined by a geodesic whose length is $d(p, q)$.*

(7.12) Corollary *Let $F_1, F_2 \colon M \to M$ be isometries of a complete, connected Riemannian manifold. Suppose that $F_1(p) = F_2(p)$ and $F_{1*} = F_{2*}$ on $T_p(M)$ for some $p \in M$. Then $F_1 = F_2$.*

Proof Let $q \in M$ and let $p(s)$, $0 \le s \le l$, be a geodesic from p to q, $p = p(0)$ and $q = p(l)$. Then $F_i(p(s))$ is a geodesic from $F_i(p)$ to $F_i(q)$, $i = 1, 2$. Since $F_1(p) = F_2(p)$ and $F_{1*}(\dot{p}(0)) = F_{2*}(\dot{p}(0))$,these geodesics coincide and

$$F_1(q) = F_1(p(l)) = F_2(p(l)) = F_2(q). \qquad ∎$$

Exercises

1. Let M be a complete Riemannian manifold and let $q \in M$. Identify $T_q(M)$ with $\boldsymbol{R}^n$, $n = \dim M$, as a manifold by choosing an orthonormal basis at q. Then $\operatorname{Exp}_q \colon T_q(M) \to M$ is a C^∞ mapping of $\boldsymbol{R}^n$ onto M with 0 mapping to q. Suppose $M = S^n$, the unit sphere with the usual metric. Prove that rank $\operatorname{Exp}_q < n$ for X_q if $\|X_q\| = k\pi$, $k = \pm 1, \pm 2, \ldots$.
2. Show that on a Riemannian manifold M which has $\boldsymbol{R}^n$ as a Riemannian covering (π is a local isometry), the rank of Exp_q is n for all $q \in M$.
3. Let M be a complete Riemannian manifold and $\pi \colon \tilde{M} \to M$ a covering. Show that there is a unique Riemannian metric on $\tilde{M}$ such that π is a local isometry and show that with this metric $\tilde{M}$ is complete.
4. Give a simple example of a Riemannian manifold diffeomorphic to $\boldsymbol{R}^n$ but such that no geodesic can be extended indefinitely.

5. Show by example that there is a Riemannian manifold on which distance between points is bounded, that is, $d(p, q) < a$, $a > 0$ fixed, but on which there is a geodesic of infinite length which does not intersect itself.
6. Show that if M is a metric space in which every bounded set is relatively compact (has compact closure), then M is complete.

8 Symmetric Riemannian Manifolds

(8.1) Definition A connected Riemannian manifold M is said to be *symmetric* if to each $p \in M$ there is associated an isometry $\sigma_p: M \to M$ which is (i) involutive (σ_p^2 is the identity), and (ii) has p as an isolated fixed point, that is, there is a neighborhood U of p in which p is the only fixed point of σ_p.

As examples we cite Euclidean n-space, in which case σ_p is reflection in p; and S^n, the unit sphere in $\boldsymbol{R}^{n+1}$, with the metric induced by $\boldsymbol{R}^{n+1}$. In the case of the sphere, σ_p is again reflection in p—for each q, $\sigma_p(q) = q'$, where q and q' are equidistant from p on a geodesic (great circle) through p (Fig. VII.13).

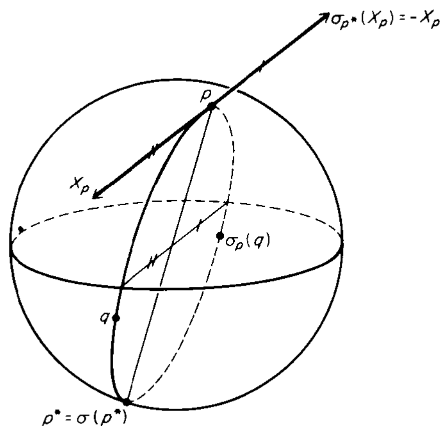


Figure VII.13

In the case of S^n we note that $\sigma_p(p) = p$ and $\sigma_p(p^*) = p^*$, p^* denoting the point antipodal to p. Thus, in general, σ_p may have other fixed points than p. Note also that the first example is a noncompact manifold and that the second is compact. A symmetric space, as we shall see, is always complete.

(8.2) Lemma *If $p \in M$, a Riemannian manifold, and σ_p is an involutive isometry with p as isolated fixed point, then $\sigma_{p*}(X_p) = -X_p$ and $\sigma_p(\operatorname{Exp} X_p) = \operatorname{Exp}(-X_p)$ for all $X_p \in T_p(M)$ for which both sides are defined.*

Proof Since σ_p^2 is the identity, the same holds for $(\sigma_{p*})^2$ on $T_p(M)$. This means that the eigenvalues of σ_{p*} on $T_p(M)$ are ± 1. However, if $+1$ is an eigenvalue, then there exists a vector $X_p \neq 0$ such that $\sigma_{p*}(X_p) = X_p$. For any isometry $F: M \to M$, $F \circ \operatorname{Exp} = \operatorname{Exp} \circ F_*$ since geodesics are preserved. This means that $\sigma_p(\operatorname{Exp} tX) = \operatorname{Exp} tX$ so the geodesic through p with initial direction X_p is pointwise fixed. This means that p is not an isolated fixed point of σ_p. Thus $+1$ is not an eigenvalue and $\sigma_{p*} = -I$, I being the identity. Since σ is an isometry, $\sigma_p(\operatorname{Exp} X_p) = \operatorname{Exp} \sigma_{p*}(X_p) = \operatorname{Exp}(-X_p)$. This means that σ_p takes each geodesic through p onto itself with direction reversed, exactly as in the two examples we have cited. ∎

The following corollary is an immediate consequence of Corollary 7.12 and the lemma:

(8.3) Corollary *Given any complete Riemannian manifold M and point $p \in M$, there can be at most one involutive isometry σ_p with p as isolated fixed point.*

(8.4) Theorem *A symmetric Riemannian manifold M is necessarily complete, and if $p, q \in M$, then there is an isometry σ_r—corresponding to some $r \in M$—such that $\sigma_r(p) = q$.*

Proof First we show that M is complete by proving that every geodesic can be extended to infinite length. Suppose $p(s)$ is a geodesic ray with s as arclength, which is defined for $0 \leq s < b$. We will show that it can be extended to a length $l > b$. Let $s_0 = \frac{3}{4}b$, and let $\sigma_{p(s_0)}$ be the symmetry in $p(s_0)$. It takes the geodesic $p(s)$ to another geodesic through $p(s_0)$ whose tangent vector at $p(s_0)$ is $-(dp/ds)_{s_0}$ and whose length is the same as that of $p(s)$. Since it has a common tangent with $p(s)$ at $p(s_0)$, it coincides with $p(s)$ on the interval $\frac{1}{2}b < s < b$ and thus extends it to a length $> \frac{3}{2}b$, which proves the statement (Fig. VII.14).

Using this it follows easily that given any $p, q \in M$ there is an isometry of M taking p to q. In fact, let r be the midpoint of a geodesic from p to q. Then the isometry σ_r takes this geodesic onto itself and carries p to q. ∎

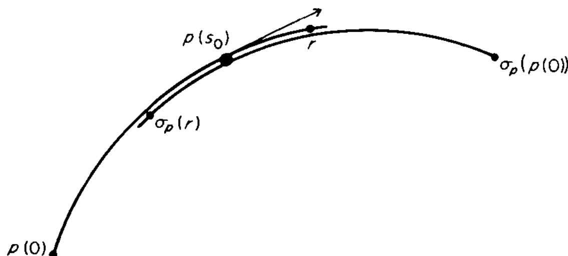


Figure VII.14

We remark here that it is easy to verify that the isometries of a Riemannian manifold M form a group $I(M)$; it is a subgroup of the group of all diffeomorphisms of M. A classical theorem due to Myers and Steenrod [1] asserts that it is a Lie group and acts differentiably on M. According to the theorem just proved it is also *transitive* when M is a symmetric space.

Before developing further the properties of symmetric spaces we prove a theorem which gives a rich collection of examples.

(8.5) Theorem *Every compact connected Lie group G is a symmetric space with respect to the bi-invariant metric.*

Proof Let $\psi: G \to G$ denote the diffeomorphism which takes each element to its inverse, $\psi(x) = x^{-1}$. This map is clearly involutive and in fact it is an isometry of G with e, the identity, as isolated fixed point. To establish this we recall that to each $X_e \in T_e(G)$ corresponds a uniquely determined one-parameter subgroup $t \mapsto g(t)$ with $\dot{g}(0) = X_e$ (Section IV.6). Since $\psi(g(t)) = g(-t)$, by the chain rule we obtain

$$\psi_*(X_e) = \psi_*(\dot{g}(0)) = \left(\frac{d}{dt}\psi(g(t))\right)_{t=0} = -\dot{g}(0) = -X_e\,.$$

This means that $\psi_{*e} = -I$, which is an orthogonal linear transformation (or isometry) of any inner product on $T_e(G)$. Let $a \in G$ be arbitrary and denote left and right translations by any $g \in G$ by L_g and R_g, respectively. We may write

$$\psi(x) = x^{-1} = (a^{-1}x)^{-1}a^{-1} = R_{a^{-1}} \circ \psi \circ L_{a^{-1}}(x).$$

Hence $\psi_{*a}: T_a(G) \to T_{a^{-1}}(G)$ may be written

$$\psi_{*a} = (R_{a^{-1}*})_e \circ \psi_{*e} \circ (L_{a^{-1}*})_a\,,$$

which is a composition of three linear mappings each of which is an isometry of the inner product determined by the bi-invariant metric ($R_{a^{-1}}$ and $L_{a^{-1}}$ induce isometries on every tangent space and ψ_{*e} is an isometry as shown above). It follows that $\psi: G \to G$ is an isometry. If we consider a normal neighborhood of e (as in Definition 6.7 with $q = e$), then by Lemma 8.2 ψ is given in local coordinates by reflection in the origin, and hence e is an isolated fixed point.

Now let $g \in G$. We define the isometry $\sigma_g: G \to G$ which has g as an isolated fixed point by $\sigma_g = L_g \circ R_g \circ \psi$, that is, $\sigma_g(x) = gx^{-1}g$. It is an isometry since R_g, L_g, and ψ are isometries, and it is easy to check that it is involutive and has g as isolated fixed point. ∎

(8.6) Example Let $G = SO(n)$ be the group of $n \times n$ orthogonal matrices of determinant $+1$. According to Example IV.6.7, the tangent space $T_e(G)$, $e = I$, the $n \times n$ identity matrix, may be identified with the skew symmetric

matrices $A = (a_{ij}) = -A'$ in the sense that $X_e = \sum_{i,j} a_{ij}(\partial/\partial x_{ij})$ is tangent at I to $SO(n)$ considered as a submanifold of $Gl(n, R) \subset \mathbf{R}^{n^2}$. The one-parameter subgroups are of the form $Z(t) = e^{tA}$. Identify $\mathfrak{g}$ with $T_e(G)$ and compute Ad B: $T_e(G) \to T_e(G)$ as follows. First one verifies from the definition of e^{tA} that

$$Be^{tA}B^{-1} = e^{tBAB^{-1}}$$

(compare Exercise IV.6.6). Since Ad(B) acting on $T_e(G)$ is just the linear map of the tangent space induced by the mapping $Z \to BZB^{-1}$ on $SO(n)$, we see that Ad(B) takes the component matrix $A = (a_{ij})$ of X_e to BAB^{-1}. Now define on $T_e(G)$ an inner product (X_e, Y_e) for $X_e = \sum a_{ij}(\partial/\partial x_{ij})$, $Y_e = \sum c_{ij}(\partial/\partial x_{ij})$ by

$$(X_e, Y_e) = \operatorname{tr} A'C = \sum_{i,j=1}^{n} a_{ij}c_{ij}.$$

It is clearly bilinear and symmetric; moreover, since

$$(X_e, X_e) = \operatorname{tr} A'A = \sum_{i,j} a_{ij}a_{ij} = \sum a_{ij}^2,$$

it is positive definite. Finally for $B \in SO(n)$

$$\begin{aligned}(\mathrm{Ad}(B)X_e, \mathrm{Ad}(B)Y_e) &= \operatorname{tr}((BAB^{-1})'BCB^{-1}) \\ &= \operatorname{tr}(BA'CB^{-1}) = \operatorname{tr} A'C = (X_e, Y_e).\end{aligned}$$

This means that this inner product determines a bi-invariant Riemannian metric on G (Lemma VI.3.4). By Theorem 8.5, G is a symmetric space with this Riemannian metric. A similar procedure may be employed to obtain the bi-invariant Riemannian metric for other compact matrix groups.

We now develop the general properties of symmetric spaces somewhat further. Let M be any symmetric Riemannian manifold and $p(t)$, $-\infty < t < \infty$, be any geodesic on M. The symmetry $\sigma_{p(t)}$ associated with any point of this geodesic maps the geodesic onto itself and reverses its sense. If c is a fixed real number, then we denote by τ_c the following composition of two such isometries, $\tau_c = \sigma_{p(c)} \circ \sigma_{p(c/2)}$. Since τ_c maps the geodesic onto itself and preserves its sense, its restriction to the geodesic must be of the form $\tau_c(p(t)) = p(t + \text{constant})$. In fact, since $\tau_c(p(0)) = \sigma_{p(c)} \circ \sigma_{p(c/2)}(p(0)) = \sigma_{p(c)}\, p(c) = p(c)$, we see that the constant is c and $\tau_c(p(t)) = p(t + c)$.

Now we consider how τ_c acts on the tangent space at a point of $p(t)$. Suppose in fact that $X_{p(0)} \in T_{p(0)}(M)$, and define a vector field $X_{p(t)}$ along $p(t)$ by the formula $X_{p(t)} = \tau_{t*} X_{p(0)}$. Let $X'_{p(t)}$ be the unique vector field satisfying $X'_{p(0)} = X_{p(0)}$ which is constant along the geodesic $p(t)$. We wish to show that these two vector fields coincide. Now for any real number t_0, $\sigma_{p(t_0)*} X'_{p(t)}$

is a parallel vector field along $p(t)$ since $\sigma_{p(t_0)}$ is an isometry. On the other hand, $\sigma_{p(t_0)*} X'_{p(t_0)} = -X'_{p(t_0)}$ since $p(t_0)$ is the fixed point of the symmetry. Because $-X'_{p(t)}$ is also a constant vector field along $p(t)$ and agrees with the field $\sigma_{p(t_0)*} X'_{p(t)}$ at one point, it must agree everywhere. Applying this argument twice we see that $\tau_{c*} X'_{p(t)} = X'_{p(t+c)}$ for all t and each constant c. Letting $t = 0$ and $c = t$ proves our assertion. We have proved the following theorem.

(8.7) Theorem *Let $p(t)$, $-\infty < t < \infty$, be a geodesic of a symmetric manifold M and τ_c the associated isometry (defined above) for each real number c. Then $\tau_c(p(t)) = p(t + c)$. If $X_{p(0)}$ is any element of $T_{p(0)}(M)$, then $X_{p(t)} = \tau_{t*} X_{p(0)}$ is the associated parallel (constant) vector field along $p(t)$, that is, as t varies $\tau_{t*}: T_{p(0)}(M) \to T_{p(t)}(M)$ is the parallel translation along the geodesic.*

(8.8) Remark Note that if $p_1 = p(c_1)$ and $p_2 = p(c_2)$ are any two points of a geodesic $p(t)$, $-\infty < t < \infty$, then by the same argument $\sigma_{p_2} \circ \sigma_{p_1}(p(t)) = p(t + 2(c_2 - c_1))$ and $(\sigma_{p_2} \circ \sigma_{p_1})_*$ maps any parallel vector field along $p(t)$ to a parallel vector field.

Theorem 8.7 will be used to prove a fact about compact Lie groups which is not at all obvious; it is given as a corollary to the following theorem. [It is because of this theorem that the notation Exp tX is used for geodesics in Riemannian manifolds.]

(8.9) Theorem *Let $M = G$, a compact, connected Lie group with the bi-invariant metric and let $X_e \in T_e(G)$. Then the unique geodesic $p(t)$ with $p(0) = e$ and $\dot{p}(0) = X_e$ is precisely the one-parameter subgroup determined by X_e. All other geodesics are left (or right) cosets of these one-parameter subgroups.*

Proof Given a geodesic $p(t)$ with $p(0) = e$, we consider the isometry $\sigma_{p(s)}\sigma_{p(0)}$ of G. By the remark above we see that this maps the geodesic onto itself with $p(t)$ being mapped to $p(t + 2s)$. But using our formula for σ_p on G together with $p(0) = e$, we have

$$\sigma_{p(s)}\sigma_{p(0)}\, p(t) = p(s)p(t)p(s).$$

[The right-hand side is the group product of $p(s)$, $p(t)$, and $p(s)$.] Thus for all t, s,

$$p(s)p(t)p(s) = p(t + 2s).$$

Using various t and mathematical induction, this gives for arbitrary s and any integer n

$$(p(s))^n = p(ns).$$

In particular, if a, b, c, d are integers with $bd \neq 0$, we have

$$p\left(\frac{a}{b}+\frac{c}{d}\right)=p\left(\frac{1}{bd}\right)^{ad+bc}=p\left(\frac{1}{bd}\right)^{ad}p\left(\frac{1}{bd}\right)^{bc}=p\left(\frac{a}{b}\right)\cdot p\left(\frac{c}{d}\right).$$

Thus for any rational numbers we have

$$p(r_1+r_2)=p(r_1)p(r_2).$$

Since $p(t)$ depends continuously on t we see that this holds for all real numbers, and thus any geodesic with $p(0)=e$ is a one-parameter subgroup. However, since there is exactly one geodesic and one such subgroup with given $\dot{p}(0)=X_e$, we see that the first sentence of the theorem is true. The second follows at once if we use the fact that either left or right translations are isometries, and hence preserve geodesics, together with the fact that a geodesic through any $g\in G$ is uniquely determined (with its parametrization) by its tangent vector at g. ∎

(8.10) Corollary *If G is a compact, connected Lie group, then any $g\in G$ lies on a one-parameter subgroup.*

Proof With the bi-invariant Riemannian metric G is a symmetric Riemannian manifold. Moreover it is complete and hence any pair of points can be joined by a geodesic. If $g\in G$, then the geodesic segment from e to g is on a one-parameter subgroup according to the theorem. ∎

(8.11) Example If $G=SO(n)$, then the geodesics, relative to the bi-invariant metric of Example 8.6 are the curves $p(t)=e^{tA}$ (A any skew symmetric matrix) and their cosets.

In the case of a group G with bi-invariant metric we can now establish a relation between the Lie derivative and the Riemannian differentiation ∇ of vector fields, which we shall need in the next chapter.

(8.12) Theorem *If X and Y are left-invariant vector fields on G and ∇ is as above, then we have*

$$\nabla_X Y=\tfrac{1}{2}[X,Y]=\tfrac{1}{2}L_X Y.$$

Proof Suppose that Z is any left-invariant vector field. Then we will compute $\nabla_{Z_e}Z$. If $g(t)$ is the uniquely determined one-parameter group with $g(0)=e$ and $\dot{g}(0)=Z_e$, then for any vector field Y we have $\nabla_{Z_e}Y=(DY_{g(t)}/dt)_{t=0}$. However, $Z_{g(t)}=dg/dt$, and $g(t)$ is a geodesic. Thus $DZ_{g(t)}/dt=(D/dt)(dg/dt)=0$ and $\nabla_{Z_e}Z=0$. Since Z and the metric are

left-invariant, it follows that $\nabla_Z Z = 0$ everywhere on G. Thus $\nabla_{X+Y}(X + Y) = 0$, from which we conclude that

$$\nabla_X Y + \nabla_Y X = 0.$$

On the other hand, we know that for any pair of vector fields a Riemannian connection satisfies the identity $\nabla_X Y - \nabla_Y X = [X, Y]$. Combining these two identities gives the lemma. ∎

Exercises

1. Let $T\colon V \to V$ be a linear transformation on a vector space over $\boldsymbol{R}$ which is involutive, that is, T^2 is the identity. Show that there is a basis of V such that the matrix of T is diagonal with diagonal elements equal to ± 1.
2. The unitary group $U(n)$ consists of all complex matrices A which satisfy the relation $AA^* = I$, $A^* = {}^t\bar{A}$, the transpose conjugate. Show that $U(n)$ can be considered as a compact subgroup of $Gl(2n, \boldsymbol{R})$ and determine a bi-invariant Riemannian metric by giving it explicitly on $T_e(U(n))$ as in Example 8.6.
3. Let G be a compact Lie group with a bi-invariant Riemannian metric and let X_e, Y_e, Z_e be vectors at the identity. Compute $R(X, Y) \cdot Z$ by extending them to left-invariant vector fields on G and using Theorem 8.12.

9 Some Examples

Except for Euclidean space itself, the examples we have given of symmetric spaces have been compact manifolds. We will consider a further example, which is not compact. To do so we must begin, in a rudimentary way at least, to develop some additional theory which will show the path toward further examples—in fact toward all examples of symmetric spaces.

As we have noted, symmetric spaces are acted upon transitively by their group of isometries. It is natural, therefore, to ask under what circumstances can one be sure that a manifold M, acted on transitively by a Lie group G, can be endowed with a Riemannian metric relative to which the transformations of M by elements of G are isometries. A sufficient condition is given by the following theorem.

(9.1) Theorem *Let G be a Lie group acting transitively on a manifold M. Then M has a Riemannian metric such that the transformation determined by each element of G is an isometry if the isotropy group H of a point $p \in M$ is a connected compact (Lie) subgroup of G.*

Proof We let $\theta: G \times M \to M$ denote the action, and for each $g \in G$, $\theta_g: M \to M$ denotes the diffeomorphic transformation of M onto itself determined by g, $\theta_g(q) = \theta(g, q)$. If $g \in H$, then $\theta_g(p) = p$ so that it induces a linear mapping $\theta_{g*}: T_p(M) \to T_p(M)$. Since $\theta_{g_1} \circ \theta_{g_2} = \theta_{g_1 g_2}$, we have $\theta_{g_1 *} \circ \theta_{g_2 *} = \theta_{g_1 g_2 *}$ so that $g \to \theta_{g*}$ is a homomorphism of H into the group of linear transformations on $T_p(M)$. From the fact that θ is C^∞ it is easily verified that this is a C^∞ homomorphism, that is, a representation of H on $T_p(M)$. Referring to Section VI.3 and, in particular, Theorem VI.3.9, we see that since H is compact and connected, there must be an invariant inner product, which we shall denote by $\Phi_p(X_p, Y_p)$ on $T_p(M)$. Now if $q \in M$, there is a $g \in G$ such that $\theta_g(q) = p$. We define $\Phi_q(X_q, Y_q)$ by

$$\Phi_q(X_q, Y_q) = \theta_g^* \Phi_p(X_q, Y_q) = \Phi_p(\theta_{g*} X_q, \theta_{g*} Y_q).$$

If $\theta_{g_1}(q) = p$ also, then $gg_1^{-1} \in H$. Hence $\theta^*_{gg_1^{-1}} \Phi_p = \Phi_p$ and

$$\theta^*_{g_1} \Phi_p = \theta^*_{g_1} \theta^*_{gg_1^{-1}} \Phi_p = \theta^*_{g_1} \circ \theta^*_{g_1^{-1}} \circ \theta^*_g \Phi_p = \theta^*_g \Phi_p .$$

It follows that Φ_q is well defined; it is positive definite since θ_g is a diffeomorphism; and it is easily verified that Φ is C^∞ and G-invariant on M. Thus Φ defines a Riemannian metric on M with respect to which each θ_g is an isometry of M. This completes the proof. ∎

In the following theorem we will continue this notation, and suppose as above that H is compact and connected and moreover that the action of G on M is effective. Then we are able to impose an additional condition which will be sufficient to ensure that M, with a metric which makes G a group of isometries, is a symmetric space. This will open the way to further examples; of which we give only one in detail.

(9.2) Theorem *With G, H, p and M as above suppose that $\alpha: G \to G$ is an involutive automorphism of G whose fixed set is H. Then the correspondence $\tilde{\alpha}(\theta(g, p)) = \theta(\alpha(g), p)$ defines an involutive isometry of M onto M with p as an isolated fixed point.*

Proof First we check that $\tilde{\alpha}$ actually defines a mapping of M onto itself. Let q be an arbitrary point of M. By transitivity there is at least one $g \in G$ such that $\theta(g, p) = q$. If g' is a second such element, then $g' = gh$ and $\alpha(g') = \alpha(g)\alpha(h) = \alpha(g)h$. Hence

$$\tilde{\alpha}(\theta(g', p)) = \theta(\alpha(g'), p) = \theta(\alpha(g)h, p) = \theta(\alpha(g), \theta(h, p)) = \theta(\alpha(g), p)$$

as required. Therefore $\tilde{\alpha}$ is defined independently of any choices. Since $\tilde{\alpha}^2$ is the identity, $\tilde{\alpha}$ is onto. Let us assume for the moment that we have proved that $\tilde{\alpha}$ is C^∞, has p as an isolated fixed point, and that $\tilde{\alpha}_*: T_p(M) \to T_p(M)$ is

$-I$, that is, $\tilde{\alpha}_*(X_p) = -X_p$. Then, clearly, $\tilde{\alpha}_*$ preserves the inner product Φ_p on $T_p(M)$. If $q \in M$, $q \neq p$, then choose $g \in G$ such that $\theta_g(p) = q$. Then

$$\tilde{\alpha}(q) = \theta(\alpha(g), p) = \theta_{\alpha(g)}(\theta_{g^{-1}}(q)).$$

Hence $\tilde{\alpha}_{*q}: T_q(M) \to T_{\tilde{\alpha}(q)}(M)$ is given by $\tilde{\alpha}_{*q} = \theta_{\alpha(g)*} \circ \theta_{g^{-1}*}$, both of which are isometries on the tangent spaces. Thus, subject to checking the other properties, $\tilde{\alpha}$ is an isometry.

In order to verify the remaining properties we need to use the fact that the natural identification of M with G/H given by the mapping $F: G/H \to M$, $F(gH) = \theta(g, p)$, is C^∞ and commutes with left translation on G/H. Thus we use Section IV.9, which was an application of Frobenius' Theorem [although, in fact, in the examples given below the facts we need here can be checked directly without relying on this general procedure]. First we recall that if $gH \in G/H$, then there is a C^∞ section S defined on a neighborhood V of gH, $S: V \to G$ with $\pi \circ S = \text{id}$ ($\pi: G \to G/H$ is the natural projection and id the identity on V). Using the diffeomorphism F, obtain a C^∞ section $\tilde{S} = S \circ F^{-1}$ on $\tilde{V} = F(V)$ into G which means a C^∞ mapping such that $\theta(\tilde{S}(q), p) = q$ for all $q \in \tilde{V}$. Every point of M is contained in the domain $\tilde{V}$ of such a section, and $\tilde{\alpha} \mid \tilde{V}$ is given by

$$\tilde{\alpha}(q) = \tilde{\alpha}(\theta(\tilde{S}(q), p)) = \theta(\alpha(\tilde{S}(q)), p),$$

which is a composition of C^∞ mappings. It follows that $\tilde{\alpha}$ is C^∞.

Finally we wish to show that $\tilde{\alpha}$ has p as an isolated fixed point and that $\tilde{\alpha}_{*p} = -I$. We use facts demonstrated in Section IV.6 concerning the exponential mapping (Definition IV.6.8)

$$\exp: T_e(G) \to G$$

(not to be confused with Exp, the exponential mapping of Riemannian manifolds). Given any $X_p \in T_e(G)$, then $\exp tX_p = g(t)$ is the one-parameter subgroup of G with $\dot{g}(0) = X_e$; and $\exp X_p = g(1)$. By Theorem IV.6.10, there is an $\varepsilon > 0$ such that an ε-ball $B_\varepsilon^n(0) \subset T_e(M)$ is mapped diffeomorphically onto a neighborhood U of e, the identity of G. Since $\alpha: G \to G$ is a Lie group automorphism with α^2 the identity, $\alpha_*: T_e(G) \to T_e(G)$ splits $T_e(G)$ into the direct sum of two subspaces $V^\pm$ of characteristic vectors belonging to the characteristic values ± 1 of α_*. Since $\alpha(\exp tX_e) = \exp t\alpha_*(X_e)$, $\alpha_*(X_e) = X_e$ if and only if $X_e \in T_e(H)$. Thus $T_e(G) = V^+ \oplus V^-$, $V^+ = T_e(H)$. $\pi: G \to G/H$ defines $\pi_*: T_e(G) \to T_{\pi(e)}(G/H)$ with $\ker \pi_* = T_e(H)$ and $\pi_* \mid V^-$ an isomorphism onto. It follows that $\pi \circ \exp$ maps a neighborhood W of $V^- \cap B_\varepsilon^n(0) \subset T_e(G)$ diffeomorphically onto a neighborhood of H in G/H. Composing with $F: G/H \to M$ gives a diffeomorphism onto an open set around p. Thus for $X_e \in W$, the mapping $X_e \to \theta(\exp X_e, p)$

is a diffeomorphism. Moreover $\tilde{\alpha}(\theta(\exp X_e, p)) = \theta(\alpha(\exp X_e), p) = \theta(\exp(-X_e), p)$. It follows that p is the only fixed point of $\tilde{\alpha}$ in this neighborhood and that $\tilde{\alpha}_*: T_p(M) \to T_p(M)$ is $-I$; each vector is taken to its negative. Combining this with what we have already established, the proof is complete. ∎

The following corollary is immediate, since each $\theta_g: M \to M$ is an isometry.

(9.3) Corollary *Under the assumptions of the theorem M is a symmetric space with involutive isometries $\sigma_p = \tilde{\alpha}$ and $\sigma_q = \theta_g \circ \tilde{\alpha} \circ \theta_{g^{-1}}$ for $q = \theta(g, p)$.*

Much of the above is to enable us to consider somewhat more complicated examples of Riemannian manifolds, of which the following is a sample.

(9.4) Example Let M be the collection of all $n \times n$, symmetric, positive definite, real matrices of determinant $+1$, and let $G = Sl(n, \boldsymbol{R})$ be the $n \times n$ matrices of determinant $+1$. Then G acts on M as follows:

$$\theta(g, s) = gsg',$$

where g' denotes the transpose of $g \in Sl(n, \boldsymbol{R})$. We will let p, the base point of the theorems above be I, the $n \times n$ identity. We then note that $H = SO(n)$ since

$$H = \{g \in Sl(n, \boldsymbol{R}) \mid \theta(g, I) = I\}$$

is given by the equivalent condition $gg' = I$, that is, $g \in SO(n)$, the group of orthogonal $n \times n$ matrices. Hence M is canonically identified with $Sl(n, \boldsymbol{R})/SO(n)$.

The automorphism α which we consider is defined by $\alpha(g) = (g^{-1})'$, the transpose of the inverse of $g \in Sl(n, \boldsymbol{R})$. Note that $\alpha(g) = g$ if and only if $g \in SO(n)$. Thus all of the conditions of the theorem are met if $Sl(n, \boldsymbol{R})$ is transitive on M. However, any positive definite, symmetric matrix q may be written in the form $q = gg' = gIg'$ where $g \in Sl(n, \boldsymbol{R})$ by standard theorems of linear algebra. From the corollary above M is a symmetric space relative to an $Sl(n, \boldsymbol{R})$ invariant metric. [Note that $\tilde{\alpha}: M \to M$ can be seen, quite directly, to be C^∞ and to have the identity $p = I$ as its *only* fixed point on M. In fact, using $q = sIs'$, we see that

$$\tilde{\alpha}(q) = \tilde{\alpha}\theta(s, I) = \theta(s'^{-1}, I) = s'^{-1}s^{-1} = (ss')^{-1} = q^{-1}.$$

Thus $\tilde{\alpha}: M \to M$ simply takes each positive definite symmetric matrix to its inverse. The only such matrix which is equal to its own inverse is the identity I.]

(9.5) Example A variant on the above, which is a particularly important case, is the following. Let $M = \{(x, y) \in \boldsymbol{R}^2 | y > 0\}$, the upper half-plane. [Note that it is covered by a single coordinate neighborhood.] We define an action of $Sl(2, \boldsymbol{R})$ on M as follows. We identify $\boldsymbol{R}^2$ with C, the complex numbers, in the usual way. Let $z = x + iy$ and let $w = u + iv$, $i = \sqrt{-1}$. When $g \in Sl(2, \boldsymbol{R})$, that is,

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad ad - bc = +1,$$

we then define $w = \theta(g, z) = (az + b)/(cz + d)$. It is not difficult to verify directly that if $y = \operatorname{Im}(z) > 0$, then $v = \operatorname{Im}(w) > 0$ and that $\theta(g_1, \theta(g_2, z)) = \theta(g_1 g_2, z)$. Moreover the Riemannian metric defined (in the local coordinates (x, y)—or $z = x + iy$—which cover M) by the matrix of components

$$(g_{ij}) = \begin{pmatrix} \dfrac{1}{y^2} & 0 \\ 0 & \dfrac{1}{y^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{(\operatorname{Im}(z))^2} & 0 \\ 0 & \dfrac{1}{(\operatorname{Im}(z))^2} \end{pmatrix}$$

is invariant under the action of $Sl(2, \boldsymbol{R})$; thus this group acts on M as a group of isometries of this metric (Exercise 4).

If we let the complex number i which corresponds to $(0, 1)$ in $\boldsymbol{R}^2$, play the role of p in the general discussion above (Theorems 9.1 and 9.2), we note the following two facts. First, the action is transitive. Given any $z_0 = u + iv$ with $v > 0$, then an element of $G = Sl(2, \boldsymbol{R})$ taking i to z_0 is

$$g = \begin{pmatrix} \sqrt{v} & \dfrac{u}{\sqrt{v}} \\ 0 & \dfrac{1}{\sqrt{v}} \end{pmatrix},$$

giving, in general,

$$\theta(g, z) = \frac{\sqrt{v}\, z + (u/\sqrt{v})}{0z + (1/\sqrt{v})} = vz + u;$$

and, when $z = i$,

$$\theta(g, i) = u + iv.$$

Second, the isotropy group of i consists of all $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in Sl(n, \boldsymbol{R})$ such that $i = (ai + b)/(ci + d)$. However, this means that $ai + b = -c + di$ or $a = d$ and $b = -c$. Since in addition $ad - bc = +1$, we have also $a^2 + b^2 = 1$; hence

$$g = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and $H = SO(2)$. It follows from our general theory in Section IV.9, that the upper half-plane with this geometry and the 2×2 positive definite matrices of Example 9.4 are equivalent both as manifolds and as homogeneous spaces with $Sl(2, \boldsymbol{R})/SO(2)$. This shows that the identification of a homogeneous space with a coset space of a Lie group as a prototype is a deeper and more interesting result than it might appear to be. In many cases rather concretely given geometric spaces can best be studied in the context of coset spaces of Lie groups.

The example we have been considering, the upper half-plane, is a realization (due to Poincaré) of the space of non-Euclidean geometry discovered by Bolyai, Lobachevskii, and Gauss. Its geometry can be studied using results of this chapter. For example, we have earlier asked the reader (Exercise 5.4) to check that the lines $x =$ constant are geodesics in this geometry. Now we propose another problem: Show that the upper halves of circles with centers on the x-axis are—when suitably parametrized—also geodesics. This is done by showing that each such circle is an image by one of the isometries of G of a vertical line. Moreover, since through a point z there is such a circle tangent to any direction, these must be *all* of the geodesics. Using this fact it is easy to see that Euclid's postulate of parallels does not hold in this geometry: There are more than one, in fact an infinite number of lines through a point z not on the line L which are parallel to L, that is, do not intersect L at any point of the upper half-plane M. The possibilities are shown in Fig. VII.15, where L_1' and L_2' indicate parallel lines (geodesics) which bound the infinite collection (faint lines) of lines L' parallel to L through z.

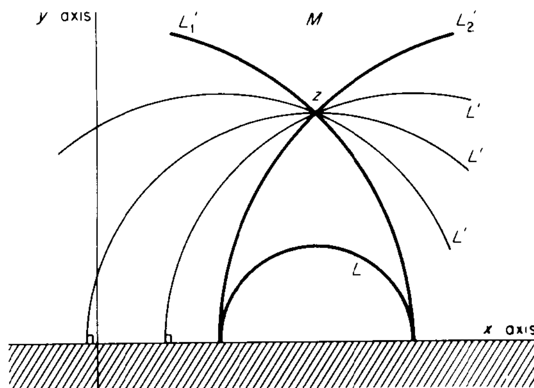


Figure VII.15

(9.6) Example As a last example of a symmetric space, we mention the Grassmann manifold $G(k, n)$ of k-planes through the origin of $\boldsymbol{E}^n$. We have noted in Section IV.9 that this is a homogeneous manifold acted on in a

natural way by $Gl(n, \boldsymbol{R})$. It is easy to see that the subgroup $SO(n, \boldsymbol{R})$ also acts transitively on the k-planes in $\boldsymbol{R}^n$. In fact, a k-plane contains an orthonormal basis $\mathbf{f}_1, \ldots, \mathbf{f}_k$ which can be completed to an orthonormal, oriented basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ of $\boldsymbol{R}^n$. Then there exists an orthogonal transformation of determinant $+1$ taking the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to this one. Hence the k-plane P_0 spanned by $\mathbf{e}_1, \ldots, \mathbf{e}_k$ is carried onto any k-plane P by at least one element of $SO(n, \boldsymbol{R})$ acting in the natural way. The isotropy group H of P_0 is $S(O(k) \times O(n-k))$, the matrices in $SO(n)$ of the form

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \qquad A \in O(k), \quad B \in O(n-k), \quad \det A \det B = +1.$$

We shall not pursue this example further except to mention that in this case α is the automorphism $\alpha\colon x \mapsto gxg^{-1}$ determined by the element

$$g = \begin{pmatrix} -I_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

of $Gl(n, \boldsymbol{R})$; then $\alpha(x) = x$ if and only if $x \in H$.

Many further details on this and other symmetric spaces may be found in Helgason [1].

Exercises

1. Prove that any positive definite, symmetric $n \times n$ matrix P is of the form $P = AA'$, where A is a nonsingular matrix of determinant $+1$. Find all possible A such that $P = AA'$ for a given P. Show that conversely AA', A an $n \times n$ matrix, is positive definite if A is nonsingular.
2. Show that if P is a positive definite symmetric matrix and $P = P^{-1}$, then $P = I$.
3. If a, b, c, d are real, compute the imaginary part of $(az + b)/(cz + d)$ and show that it has the same sign as $\operatorname{Im}(z)$.
4. Prove that $Sl(2, \boldsymbol{R})$, acting as in Example 9.5, leaves the Riemannian metric given there invariant, that is, each transformation of this type is an isometry of the upper half-plane. Show that under these isometries circles with center on the x-axis go into circles of the same type, or vertical lines.
5. Show that the mapping $A \to e^A$ maps the symmetric $n \times n$ matrices onto the positive definite symmetric matrices and is one-to-one and onto. Use this to show that curves through I given by $P(t) = e^{tA}$, A symmetric, $n \times n$, allow us to identify $T_I(M)$, the tangent space at I to the manifold M of positive definite symmetric matrices of determinant $+1$ with the symmetric $n \times n$ matrices of trace zero. (See Example 9.4.)

6. Show that $(A, B) = \mathrm{tr}(AB)$ defines an inner product on the space of symmetric $n \times n$ matrices. Using Exercise 5, show how this may be used to define a Riemannian metric on M which is invariant under the action θ: $Sl(n, \boldsymbol{R}) \times M \to M$ of Example 9.4.
7. Verify that α defined in Example 9.6 is an involutive automorphism of $SO(n)$ leaving fixed the subgroup H.
8. Show that in Example 9.5 there is a unique geodesic segment joining any pair of points p, q and that its length is the distance $d(p, q)$ as defined in Section V.3.
9. Show that the mapping $A \to e^A$ of $\mathscr{M}_2(\boldsymbol{R}) \to Sl(2, \boldsymbol{R})$ is not onto.

Notes

The proper generalization of differentiation from Euclidean manifolds to Riemannian manifolds was difficult to discover and came long after the work of Gauss and Riemann. The notion of parallel displacement of vector fields along curves generally attributed to Levi-Civita [1] furnished the basic idea from which the theory developed. The use of the operator ∇_X and its axiomatization are much more recent and are due to Koszul [1]. Many of the references contain a more complete theory of connections and of differentiation on manifolds which does not depend on a Riemannian metric—and hence is not unique.

The curvature, which is introduced so briefly here, is in some sense the obstacle to differentiating exactly as in Euclidean space, for there parallel displacement of a vector field along a curve from p to q is independent of the path chosen. In the Riemannian case, however, it is not the same along every curve. Thus parallel displacement of $T_p(M)$ along a closed curve (loop) at p yields the identity transformation of $T_p(M)$ if we are in Euclidean space, and a linear transformation related to the curvature operator otherwise.

Once differentiation is successfully generalized, one can begin the study of Riemannian geometry itself. We began this with the study of geodesics and, in the next chapter we shall go on to a brief study of curvature. Of course, we do not dig very deeply; both geodesics and curvature as well as their interrelations are the basis for an enormous amount of interesting research. Indeed, geodesics and curvature are the basic objects of study in Riemannian Geometry. The reader should consult Bishop and Crittenden [1], Kobayashi and Nomizu [1], Milnor [1], and other books listed in the references for further work and extensive bibliographies of these topics.

Symmetric spaces are among the most important examples of Riemannian manifolds. In particular, they include as special cases all of the classical geometries: Euclidean, non-Euclidean, projective (real and complex); they were first studied by Elie Cartan in several important papers and were completely classified by him. They are all homogeneous spaces of Lie groups and indeed their classification is closely related to the classification of real, semisimple Lie algebras, also carried out by Cartan. The reader should consult Chern and Chevalley [1] and Helgason [1] for more details. Some further examples are discussed in Chapter VIII.

VIII CURVATURE

This chapter continues our brief introduction to Riemannian geometry by defining and interpreting the concept of curvature. This is the most important invariant of a Riemannian metric on a manifold and completely determines the local geometry. Its definition requires the operation of differentiation developed in Chapter VII.

We begin with a brief exposition of the geometry of surfaces in E^3, that is, two-dimensional submanifolds imbedded in ordinary Euclidean space; E^3 comes equipped with a Riemannian metric and thus induces one on the surface. Using differentiation of vector fields along curves on the surface M we are able to define a symmetric bilinear form (covariant tensor of order 2) on M which is related to the shape of the surface, and a corresponding symmetric (self-adjoint) operator on the tangent spaces to M whose trace and determinant are the mean and Gaussian curvatures. The latter, denoted by K, is of fundamental importance because of the profound discovery of Gauss that it is unaltered by modifications of the manner in which M is imbedded so long as lengths of curves (and hence the Riemannian metric) are unaltered. This is not proved until Section 4; but in Section 3 we deduce the basic symmetry properties of the Riemannian curvature $R(X, Y, Z, W)$, a covariant tensor of order 4 which was defined in Section VII.4. These properties are then used in Section 4 to prove Gauss's theorem and to determine the relation of the Gaussian curvature and the Riemannian curvature. This involves the important idea of *sectional curvature* in an arbitrary Riemannian manifold, which is defined and discussed in Section 4, using the various symmetries of $R(X, Y, Z, W)$.

In Section 5 we extend the differentiation process previously defined for vector fields to arbitrary covariant tensor fields and use it to define the notion of parallel vector fields. We state (without proof) the local characterization of symmetric spaces as Riemannian manifolds whose

curvature tensor is parallel. This includes, in particular, the manifolds whose curvature is constant on all sections, called manifolds of constant curvature. The last section is devoted to a discussion of these Riemannian manifolds; they include all of the classical geometries: Euclidean, hyperbolic or non-Euclidean, and elliptic (real projective space).

1 The Geometry of Surfaces in E^3

In this section we use our earlier definitions of curvature for curves in Euclidean three-dimensional space to obtain various quantities which measure the shape of a surface M near each of its points. All of these will depend for their definition on the fact that the surface M lies in Euclidean space; however they will be independent of the coordinates used both on M and on E^3, as will be seen from their definition. Since all properties are local in character, we suppose that M is an *imbedded* surface of which we consider only a portion covered by a single coordinate neighborhood U, φ with $W = \varphi(U)$ a connected open subset of R^2, the uv-plane. Thus $p \in U \subset M$ has coordinates $(u(p), v(p)) = \varphi(p)$; and, taking the Euclidean three-dimensional space with a fixed Cartesian coordinate system, that is, identifying E^3 with R^3, the imbedding or parameter mapping $\varphi^{-1}: W \to U \subset R^3$ is given by $x^i = f^i(u, v)$, $i = 1, 2, 3$. Let $E_1 = \varphi_*^{-1}(\partial/\partial u)$ and $E_2 = \varphi_*^{-1}(\partial/\partial v)$ denote the coordinate frames and suppose further that M is orientable and oriented with U, φ giving the orientation. This is an important condition on M since we are then able to define, without ambiguity, the *unit normal* vector field N to M; it is the unique unit vector at each $p \in M$ which is orthogonal to $T_p(M) \subset T_p(R^3)$ and so chosen that E_1, E_2, N form a frame at p with the same orientation as $\partial/\partial x^1, \partial/\partial x^2, \partial/\partial x^3$ the standard orthonormal frame of R^3. Length and orthogonality are defined in terms of the inner product (X, Y) of Euclidean space which, of course, induces a Riemannian metric on M by restriction. We shall study the shape of M at $p \in M$ by means of the derivative of N in various directions tangent to M at p.

In fact, using the ideas developed in Sections VII.1 and VII.2, let $p(t)$ be any differentiable curve on M with $p(0) = p$ and $\dot{p}(0) = X_p \in T_p(M)$. Restricting N to $p(t)$ gives a vector field $N(t) = N_{p(t)}$ along $p(t)$ which may be differentiated in R^3 as a vector field along a space curve, giving a derivative dN/dt which is itself a vector field along $p(t)$. Applying (VII.1.4c) and using $(N, N) = 1$, we have

$$0 = \frac{d}{dt}(N, N) = 2\left(\frac{dN}{dt}, N\right).$$

This means that dN/dt is orthogonal to $N(t)$ at each point $p(t)$ and hence is tangent to M, that is, $dN/dt \in T_{p(t)}(M)$ (see Fig. VIII.1).

If we restrict our attention to a fixed point $p \in M$, and consider various curves through it with $p(0) = p$ and tangent vector $X_p = \dot{p}(0)$, then we have the following result.

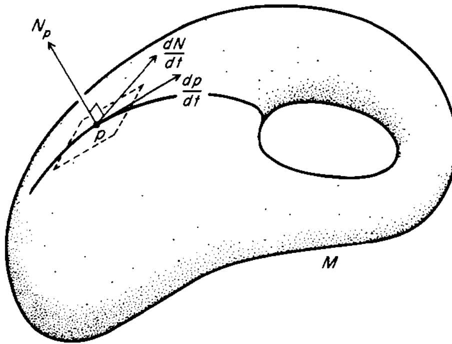


Figure VIII.1

(1.1) Theorem *The vector $(dN/dt)_{t=0}$ depends only on X_p and not on the curve $p(t)$ chosen. Let $S(X_p) = -(dN/dt)_{t=0}$. Then $X_p \to S(X_p)$ is a linear map of $T_p(M) \to T_p(M)$.*

Proof Let $X_p = aE_{1p} + bE_{2p}$ be an arbitrary element of $T_p(M)$ written as a linear combination of the coordinate frame E_{1p}, E_{2p} of the coordinate neighborhood U, φ containing p. Let

$$p(t) = \big(f^1(u(t), v(t)), f^2(u(t), v(t)), f^3(u(t), v(t))\big)$$

be any differentiable curve with $p(0) = p$, $\dot{p}(0) = X_p$ and suppose $p(0)$ has coordinates $u_0 = u(0)$ and $v_0 = v(0)$. Since $\dot{p}(0) = X_p$, we have $\dot{p}(0) = aE_{1p} + bE_{2p}$, that is, $\dot{u}(0) = a$ and $\dot{v}(0) = b$. We denote by $n^i(u, v)$ the components of N on U relative to the standard frames in $\boldsymbol{R}^3$

$$N = n^1(u, v)\frac{\partial}{\partial x^1} + n^2(u, v)\frac{\partial}{\partial x^2} + n^3(u, v)\frac{\partial}{\partial x^3}.$$

Then, along the curve

$$N(t) = \sum_{i=1}^{3} n^i(u(t), v(t))\frac{\partial}{\partial x^i}$$

and

$$\left(\frac{dN}{dt}\right)_0 = \sum_{i=1}^{3}\left[\left(\frac{\partial n^i}{\partial u}\right)_{\varphi(p)}\dot{u}(0) + \left(\frac{\partial n^i}{\partial v}\right)_{\varphi(p)}\dot{v}(0)\right]\frac{\partial}{\partial x^i}$$
$$= a\left(\sum_{i=1}^{3}\left(\frac{\partial n^i}{\partial u}\right)_{\varphi(p)}\frac{\partial}{\partial x^i}\right) + b\left(\sum_{i=1}^{3}\left(\frac{\partial n^i}{\partial v}\right)_{\varphi(p)}\frac{\partial}{\partial x^i}\right).$$

This shows that $S(X_p)$ depends linearly on the components of X_p, and since $(dN/dt)_{t=0}$ lies in $T_p(M)$, we have $S\colon T_p(M) \to T_p(M)$ is a linear map. Moreover only the values $(u(0), v(0))$, the coordinates of p, and $\dot{u}(0)$, $\dot{v}(0)$, the components of $\dot{p}(0) = X_p$, appear in the formula. Thus $(dN/dt)_0$ depends on p and X_p and not on the curve used in the calculation. ∎

(1.2) Remark The linear map $S: T_p(M) \to T_p(M)$ given at each $p \in M$ does not depend on the choice of coordinate system U, φ on M nor on the Cartesian coordinate system used in Euclidean space. This is because N is defined using only the orientations of M and Euclidean space and the inner product of the Euclidean space; the differentiation depends only on the existence of parallel orthonormal frames in Euclidean space. Thus N, dN/dt, and S are independent of coordinates and involve only the geometry of M as an imbedded surface in Euclidean space. The operator S has been appropriately called the *shape operator*, see O'Neill [1], in whose work the reader may find a detailed discussion.

By way of examples, suppose M is the xy-plane. Then $N = E_3$, a constant vector, so that $S(X_p) = 0$. On the other hand if M is a sphere of radius R, the unit normal N at $(x^1, x^2, x^3) \in M$ is given by

$$N = \frac{x^1}{R}\frac{\partial}{\partial x^1} + \frac{x^2}{R}\frac{\partial}{\partial x^2} + \frac{x^3}{R}\frac{\partial}{\partial x^3}.$$

If we move in any direction tangent to the sphere along a great circle curve, parametrized by arclength so that $\|X_p\| = 1$, then $S(X_p) = -dN/ds = -(1/R)X_p$. Further examples will be considered presently.

We may use the linear map $S: T_p(M) \to T_p(M)$—more accurately denoted S_p—which we have determined at each $p \in M$ to define a C^∞ covariant tensor field on M, assuming (as we shall henceforth) that M is a C^∞ submanifold. We follow a standard procedure from linear algebra: Let $S: V \to V$ be a linear operator on a vector space V with inner product (X, Y). Then the formula

$$\Psi(X, Y) = (S(X), Y)$$

defines a bilinear form, or covariant tensor of order 2, on V. The form Ψ is symmetric if and only if

$$(S(X), Y) = (X, S(Y))$$

holds for all $X, Y \in V$; S is then called *symmetric* or *self-adjoint*. For the linear algebra involved the reader is referred to Exercise 9 and to Hoffman and Kunze [1].

(1.3) Theorem *$S(X)$ is a symmetric operator on the tangent space $T_p(M)$ for each $p \in M$ and $\Psi(X, Y)$ is a symmetric covariant tensor of order 2. The components of S and Ψ are C^∞ if M is a C^∞ submanifold.*

Proof In order to prove these statements we compute the components of $\Psi(X, Y)$. As above U, φ is a coordinate neighborhood and $\varphi^{-1}: W \to U \subset M$ is the corresponding parametrization. The components

of $\Psi(X, Y)$ relative to the coordinate frames $E_1 = \varphi_*^{-1}(\partial/\partial u)$ and $E_2 = \varphi_*^{-1}(\partial/\partial v)$ are given by the standard formulas below, in which we use $\partial N/\partial u$ and $\partial N/\partial v$ to denote the derivatives of N along the coordinate curves on M obtained by holding one corrdinate fixed and allowing the other to vary (as parameter along the curve):

$$\Psi(E_1, E_1) = (S(E_1), E_1) = -\left(\frac{\partial N}{\partial u}, E_1\right),$$

$$\Psi(E_1, E_2) = (S(E_1), E_2) = -\left(\frac{\partial N}{\partial u}, E_2\right),$$

$$\Psi(E_2, E_1) = (S(E_2), E_1) = -\left(\frac{\partial N}{\partial u}, E_1\right),$$

$$\Psi(E_2, E_2) = (S(E_2), E_2) = -\left(\frac{\partial N}{\partial u}, E_2\right),$$

If we denote by $X = X(u, v)$ the position vector from 0 to $\varphi^{-1}(u, v)$

$$X = f^1(u, v)\frac{\partial}{\partial x^1} + f^2(u, v)\frac{\partial}{\partial x^2} + f^3(u, v)\frac{\partial}{\partial x^3},$$

then $X_u = E_1$ and $X_v = E_2$ are just the vectors whose components are the corresponding derivatives of the components of X with respect to u and v, that is, $X_u = \partial X/\partial u = E_1$ and $X_v = \partial X/\partial v = E_2$. Remembering that $(N, X_u) = 0 = (N, X_v)$ and differentiating, we obtain

$$-\left(\frac{\partial N}{\partial u}, X_u\right) = (N, X_{uu}) = \sum n_i \frac{\partial^2 f^i}{\partial u^2},$$

$$-\left(\frac{\partial N}{\partial v}, X_u\right) = (N, X_{vu}) = \sum n_i \frac{\partial^2 f^i}{\partial v\, \partial u} = (N, X_{uv}) = -\left(\frac{\partial N}{\partial u}, X_v\right),$$

$$-\left(\frac{\partial N}{\partial v}, X_v\right) = (N, X_{vv}) = \sum n_i \frac{\partial^2 f^i}{\partial v^2}.$$

These computations show that the components of Ψ, and hence of S, are C^∞ if M is. The second of these relations shows that $\Psi(X, Y) = \Psi(Y, X)$ so the tensor Ψ is symmetric. The 2×2 matrix $(l_{ij}) = (\Psi(E_i, E_j))$ of its components will often be written

$$\begin{pmatrix} l & m \\ m & n \end{pmatrix},$$

where $l = (N, X_{uu}) = l_{11}$, $m = (N, X_{uv}) = l_{12} = l_{21}$, and

$$n \doteq (N, X_{vv}) = l_{22}.$$

The bilinear form $\Psi(X, Y)$ is called the *second fundamental form* of the surface M, and the inner product (X, Y) the *first fundamental form.* Although the components of the Riemannian metric (X, Y) are denoted by g_{ij} for the general Riemannian case, one often uses E, F, G in the classical case of a surface M in Euclidean space. Thus

$$g_{11} = E = (X_u, X_u),$$
$$g_{12} = F = (X_u, X_v) = (X_v, X_u) = F = g_{21},$$
$$g_{22} = G = (X_v, X_v).$$

∎

(1.4) Remark It is a classical theorem of differential geometry (which we shall not prove) that two surfaces M_1 and M_2 in $\boldsymbol{R}^3$ are *congruent* if and only if they correspond in such fashion that at corresponding points both fundamental forms agree (O'Neill [1, p. 297; Stoker [1, p. 138]). Of course the "only if" part is immediate from the definitions. This fact shows the importance of these two forms in the geometry of the surface.

The Principal Curvatures at a Point of a Surface

Having once proved that $S(X)$ is a self-adjoint linear operator on $T_p(M)$ at each $p \in M$, we can use standard theorems of linear algebra, together with what we have learned of curves in space, to study the geometry of M.

(1.5) Theorem *At each $p \in M$ the characteristic values of the linear transformation S are real numbers k_1 and k_2, $k_1 \geq k_2$. If $k_1 \neq k_2$, then the characteristic vectors belonging to them are orthogonal; if $k_1 = k_2 = k$ at p, then $S(X_p) = kX_p$ for every vector X_p in $T_p(M)$. The numbers k_1 and k_2 are the maximum and minimum values of $\Psi(X_p, X_p) = (S(X_p), X_p)$ over all unit vectors $X_p \in T_p(M)$.*

Proof These statements are taken directly from theorems of linear algebra, but we shall sketch a proof for the case $k_1 \neq k_2$, leaving the case $k_1 = k_2$ to Exercises 7 and 8. All vectors are elements of $T_p(M)$, p fixed, in the following proof. We suppose $k_1 > k_2$ are the characteristic values, which are real since S is self-adjoint (Exercise 7), and we let F_1, F_2 be characteristic vectors of unit length corresponding to k_1, k_2. We have

$$k_1(F_1, F_2) = (S(F_1), F_2) = (F_1, S(F_2)) = k_2(F_1, F_2),$$

which implies $(F_1, F_2) = 0$ when $k_1 \neq k_2$, as assumed. Replacing F_2 by $-F_2$ if necessary, we may suppose F_1, F_2 is an orthonormal basis with the same orientation as $T_p(M)$.

Next we show that k_1 and k_2 are the maximum and minimum values of $(S(X_p), X_p)$ for unit vectors X_p. Any unit vector $X_p \in T_p(M)$ may be written

$X_p = \cos\theta\, F_1 + \sin\theta\, F_2$. Let $k(\theta)$ denote $\big(S(X_p), X_p\big) = \Psi(X_p, X_p)$. Since F_1, F_2 is an oriented, orthonormal frame, we have

$$(*) \qquad k(\theta) = k_1 \cos^2\theta + k_2 \sin^2\theta \qquad \text{(Euler's formula).}$$

Differentiating gives

$$\frac{dk}{d\theta} = 2(k_2 - k_1)\sin\theta\cos\theta.$$

Hence the extrema of $k(\theta)$ occur when $\theta = 0, \frac{1}{2}\pi, \pi$, or $\frac{3}{2}\pi$; in other words, when $X_p = \pm F_1$ or $\pm F_2$ so that k_1 and k_2 are maximum and minimum values of $\big(S(X_p), X_p\big)$ as claimed. ∎

We remark that k_1 and k_2 are the maximum and minimum of the expression $\Psi(X_p, X_p)/(X_p, X_p)$ over *all* $X_p \neq 0$ in $T_p(M)$. The points p at which $k_1 = k_2$ are called *umbilical points* of M if $k_1 \neq 0$ and *planar points* otherwise. Note that a sphere of radius R consists entirely of umbilical points with $k_1 = 1/R = k_2$. Similarly, if M is a plane, every point is planar with $k_1 = 0 = k_2$.

We shall now interpret $k(\theta) = \Psi(X_p, X_p)$ geometrically. Let p be a point of M and X_p a *unit* tangent vector at p; X_p and N_p determine a plane on which we may take p as origin and X_p, N_p as unit vectors along the axes (in that order), thus giving a coordinate system and orientation on the plane (see Fig. VIII.2). The plane intersects M along a curve which, of course, lies

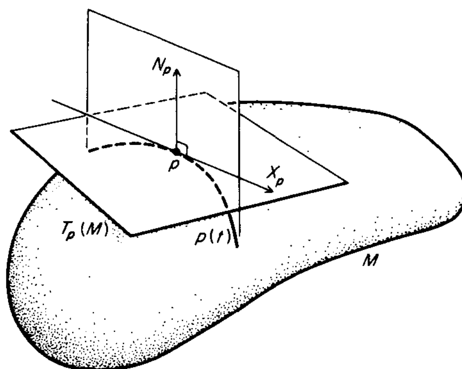


Figure VIII.2

on M and on the plane, and passes through p. It is called the *normal section* at p determined by X_p; there is clearly such a curve for each X_p. The vector N_p is the normal to the curve at p and X_p is its unit tangent vector. Writing this curve as $p(t)$ with $p(0) = p$ and with arclength as parameter, we have $\dot{p}(t) = dp/dt$ a unit vector for every t so that $\dot{p}(0) = X_p$. Differentiating

$(N, dp/dt) = 0$ along the curve, we find that $(dN/dt, dp/dt) = -(N, d^2p/dt^2) = -\tilde{k}$, the curvature of the plane curve $p(t)$ as defined at the end of Section VII.1. In particular, at $p = p(0)$, $(dN/dt, X_p) = -(S(X_p), X_p)$. Thus with $X_p = \cos\theta\, F_1 + \sin\theta\, F_2$ as above, we find that $k(\theta) = \tilde{k}$ is the curvature of the normal section determined by X_p. For this reason $k(\theta)$ is called the *normal curvature* (of the section determined by X_p); and k_1 and k_2, the maximum and minimum of $k(\theta)$, are called *principal curvatures* at p and the corresponding unit vectors F_{1p}, F_{2p} (chosen to conform to the orientation) are called *principal directions* at p.

To study the surface at p we will now choose an xyz-coordinate system in Euclidean space so that the origin is at p, $T_p(M)$ is the xy-plane, and the principal directions F_{1p}, F_{2p} and unit normal N_p at p are $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$, unit vectors on the x-, y-, z-axes, respectively. Let $x = u$, $y = v$, and $z = f(u, v)$ be the (parametric) equation of the surface. Then we may identify the xy- and uv-planes and assume that the parameter mapping φ^{-1} takes some open set W on the xy-plane onto an open set U on M. The conditions then imply

$$f(0, 0) = 0 \qquad \text{and} \qquad f_x(0, 0) = 0 = f_y(0, 0).$$

If we compute the components of the first fundamental form at p, we obtain $E = 1 = G$ and $F = 0$. For the second fundamental form, recall that $\varphi^{-1}: (x, y) \to (x, y, f(x, y))$ is the parametric representation of M and thus at p, $l = (\partial/\partial z, f_{xx}\, \partial/\partial z) = f_{xx}$, $m = (\partial/\partial z, f_{xy}\, \partial/\partial z) = f_{xy}$, and

$$n = (\partial/\partial z, f_{yy}\, \partial/\partial z) = f_{yy}.$$

Now the fact that we have chosen coordinate axes so that $\partial/\partial x$ and $\partial/\partial y$ are principal directions tells us that $m = 0$ and $l = k_1, n = k_2$. Thus we have

$$k(\theta) = f_{xx} \cos^2\theta + f_{yy} \sin^2\theta \qquad \text{at} \quad x = 0, \quad y = 0.$$

Let $f(x, y)$ be expanded in Taylor series at $(0, 0)$. Then

$$z = f(x, y) = f_{xx}(0, 0)x^2 + f_{yy}(0, 0)y^2 + R_2,$$

where R_2 contains terms of higher order. Let $f_{xx}(0, 0) = a$ and $f_{yy}(0, 0) = b$. Then we see that the normal sections of $z = ax^2 + by^2$ have the same sectional curvatures at p as does the given surface. Therefore the quadric surfaces must give typical examples.

(1.6) Example $z = ax^2 + by^2$, $ab > 0$ (see Fig. VIII.3a). This is an elliptic paraboloid; the principal curvatures are a and b. If both are positive, it lies above the xy-plane; if both are negative, it lies below. In either case when k_1 and k_2 have the same sign, the surface is (locally) on one side of $T_p(M)$.

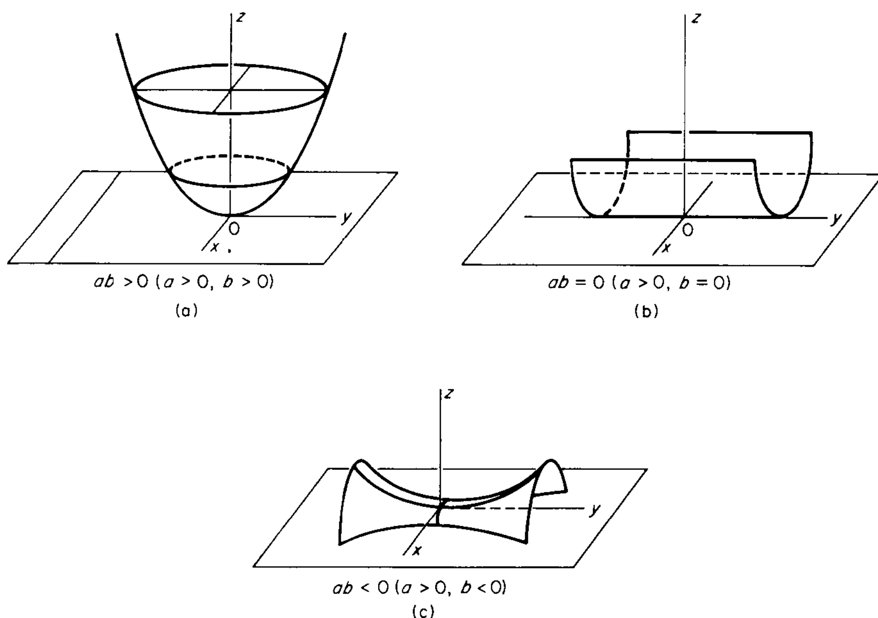


Figure VIII.3

(1.7) Example $z = ax^2 + by^2$, $ab = 0$ (see Fig. VIII.3b). If both are zero, we have the xy-plane as our surface; if one, say $b = 0$, then we have a parabolic cylinder which is above the xy-plane if $a > 0$.

(1.8) Example $z = ax^2 + by^2$, $ab < 0$ (see Fig. VIII.3c). In this case we have a hyperbolic paraboloid or saddle surface with the xy-plane tangent at the saddle. Suppose, for example, $a = 1$ and $b = -1$. Then by Euler's formula (*) $k(\theta) = \cos^2 \theta - \sin^2 \theta$ and hence $k(\theta)$ varies from $+1$ to -1 and is zero at $\pm\pi/4$, $\pm 3\pi/4$. When $k_1 > 0$ and $k_2 < 0$, then the surface must have points (locally) on both sides of $T_p(M)$.

In these exercises we follow the notation of the text.

Exercises

1. Show that if $p \in M$ is not an umbilical or planar point, then there exist coordinates U, φ on a neighborhood of p such that the curves $u =$ constant and $v =$ constant are tangent at each point to the principal directions. [These curves are called *lines of curvatures*.]
2. Let M be the surface obtained by revolving a curve $z = f(x)$ around the z-axis. Show that the lines of curvature (Exercise 1) are the circles

$z =$ constant on M and the curves obtained by intersection of M with planes containing the z-axis. Determine the umbilical points.

3. Let U, φ be coordinates such that $u =$ constant and $v =$ constant are lines of curvature. Show that in the first fundamental form the component F is zero in these coordinates and that the principal curvatures are l/E and n/G.
4. A direction X_p at $p \in M$ such that $\Psi(X_p, X_p) = (S(X_p), X_p) = 0$ is called an *asymptotic direction*. Show by example that there may be two, one, or no asymptotic directions at a point of M. Find the asymptotic directions at each point of a hyperboloid of one sheet, $(x^2/a^2) + (y^2/b^2) - (z^2/c^2) = 1$.
5. Show that if there are two distinct asymptotic directions at $p \in M$, then there exist coordinates U, φ around p such that $u =$ constant and $v =$ constant are everywhere tangent to asymptotic directions (they are *asymptotic lines*). Find the asymptotic lines for a hyperboloid of one sheet.
6. For a surface of the form $z = f(x, y)$ find the components of the first and second fundamental forms and the directions of the lines of curvature and asymptotic lines.

In the following exercises assume V is a Euclidean vector space with inner product $(\mathbf{v}, \mathbf{w})$ and that $S\colon V \to V$ is a linear operator.

7. Suppose that S is self-adjoint. Show that its matrix relative to an orthonormal basis is symmetric. If $\dim V = 2$, use this to show that the discriminant of its characteristic polynomial is not negative, so that the characteristic roots are real.
8. When S is self-adjoint and $\dim V = 2$, show that if $\mathbf{u} \neq 0$ is a characteristic vector of S, then so is any $\mathbf{v}$ orthogonal to $\mathbf{u}$. Prove, using Exercise 7, that if the discriminant is zero, then S is a scalar multiple of the identity transformation.
9. Show that the correspondence $S \leftrightarrow (S(\mathbf{v}), \mathbf{w})$ is an isomorphism between the space of linear operators on V and the space of bilinear forms on V.
10. Give a precise definition of what would be meant by a C^∞ field of linear operators on a manifold M. If M is Riemannian, show that the collection of such fields is isomorphic in a natural way to $\mathcal{T}^2(M)$.

2 The Gaussian and Mean Curvatures of a Surface

The negative of the trace and determinant of any matrix of the linear transformation S defined in Section 1 are the coefficients of the characteristic polynomial of S and are important invariants. The determinant is

$K = k_1 k_2$, the product of the characteristic values; it is called the *Gaussian curvature* of the surface. The trace is $k_1 + k_2$, the sum of the characteristic values; and $H = \frac{1}{2}(k_1 + k_2)$ is called the *mean curvature* of the surface. These quantities may be computed directly from the components of the fundamental forms, using any parametrization of the surface. This we now proceed to do.

(2.1) Theorem

$$K = \frac{ln - m^2}{EG - F^2} \qquad \textit{and} \qquad H = \frac{1}{2}\frac{Gl - 2Fm + En}{EG - F^2}.$$

Proof Together,

$$S(X_u) = aX_u + bX_v, \qquad S(X_v) = cX_u + dX_v$$

give the components of the operator S in terms of the coordinate frames $E_1 = X_u$ and $E_2 = X_v$ naturally given by the parametrization of M near p, that is, on the coordinate neighborhood U, φ. Thus we may write

$$K = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \qquad \text{and} \qquad 2H = a + d.$$

In terms of X_u, X_v we have

$$K\,(X_u \times X_v) = S(X_u) \times S(X_v)$$

and

$$2H(X_u \times X_v) = S(X_u) \times X_v + X_u \times S(X_v),$$

where $\times$ denotes the cross product of vectors in three-dimensional Euclidean space.

Now note that $(X_u \times X_v, X_u \times X_v) = \|X_u \times X_v\|^2 = EG - F^2$ and use the fact that for any vectors X, Y, U, V of $\boldsymbol{R}^3$ we have the Lagrange identities

$$((X \times Y), (U \times V)) = \begin{vmatrix} (X, U) & (X, V) \\ (Y, U) & (Y, V) \end{vmatrix}.$$

Then we obtain the formulas for K and H by taking the scalar product on both sides with $X_u \times X_v$ in each of the equations above. ∎

Since the Gaussian curvature K is the product of the principal curvatures k_1 and k_2, we see that $K > 0$ at p if both k_1 and k_2 are different from zero and have the same sign. This means that either $k_1 > 0$ and $k_2 > 0$ and the curve of each normal section curves toward the normal so that the surface lies entirely on the same side of the tangent plane as the normal N_p sufficiently near the point p, or $k_1 < 0$ and $k_2 < 0$ and each curve goes away

from the normal so that the surface (near p) lies entirely on the opposite side to N_p. Equivalently, introducing local coordinates in $\boldsymbol{R}^3$ as in Examples 1.6–1.8, $K > 0$ if and only if the function $z = f(x, y)$ has a strict relative extremum at the point.

On the other hand, if $K < 0$, then k_1 and k_2 are different from zero and have opposite signs. This means that the surface is like a saddle surface: some normal sections are concave toward the normal N and some concave away from it.

When $k = 0$ one of the principal curvatures must be zero and then little can be said. Two examples, in addition to the plane, are $z = (x^2 + y^2)^2$, which is obtained by revolving $z = x^4$ around the z-axis, and $z = x(x^2 - 3y^2)$, the so-called *monkey saddle*, which is similar to the usual saddle surface except that there are three valleys running down from the pass: two for the monkey's legs and one for its tail (Fig. VIII.4).

The mean curvature will be of less concern to us than the Gaussian curvature for reasons that will appear later. Surfaces for which the mean curvature vanishes are of special interest, however. They are *minimal surfaces*; they are like the surfaces formed by a soap film stretched over a wire frame (Fig. VIII.5). They have the defining property of being surfaces of

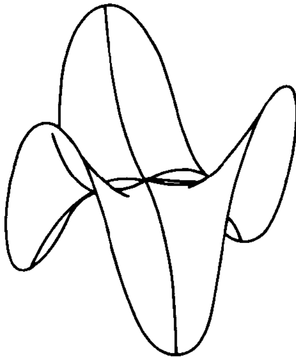


Figure VIII.4

Monkey saddle.

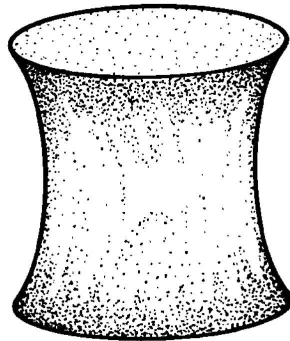


Figure VIII.5

Minimal surface.

minimal area among all surfaces with a given boundary (the wire frame). Thus, in a sense, they generalize the geodesics—curves of minimal length joining two fixed points. Like the equation of geodesics, the vanishing of the mean curvature guarantees the property of minimality only in a local sense.

(2.2) Example We consider a torus; then intuitively we can see that the two circles running around the torus which are the points of contact with the two parallel tangent planes orthogonal to its axis divide the torus into an

inner portion on which $K < 0$ and an outer portion at which $K > 0$. Along the two circles $K = 0$, since along these circles the normal vector remains parallel to the z-axis (Fig. VIII.6).

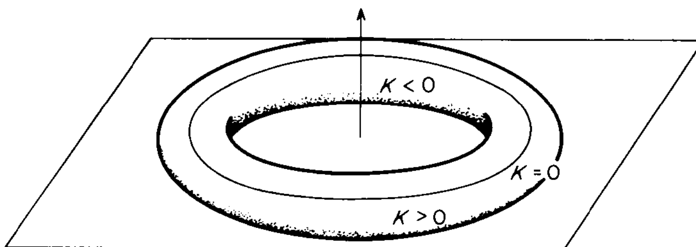


Figure VIII.6

(2.3) Example Let $(u, v) \to (u, v, uv)$ parametrize the saddle surface $z = xy$. Then $X_u = (\partial/\partial x^1) + v(\partial/\partial x^3)$ and $X_v = (\partial/\partial x^2) + u(\partial/\partial x^3)$ from which

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} 1 + v^2 & uv \\ uv & 1 + u^2 \end{pmatrix}.$$

Moreover $\lambda N = (-v, -u, 1)$ with $\lambda = (1 + u^2 + v^2)^{1/2}$ and $X_{uu} = 0 = X_{vv}$, $X_{vu} = \partial/\partial x^3$. It follows that

$$\begin{pmatrix} l & m \\ m & n \end{pmatrix} = \begin{pmatrix} 0 & \dfrac{1}{\lambda} \\ \dfrac{1}{\lambda} & 0 \end{pmatrix}.$$

From this we obtain

$$K = -\left(\frac{1}{\lambda^4}\right), \qquad H = \frac{-uv}{\lambda^3}.$$

The Theorema Egregium of Gauss

The entire subject of differential geometry was influenced by a very profound discovery of Gauss which may be stated as follows:

(2.4) Theorem (Gauss) *Let M_1 and M_2 be two surfaces in Euclidean space and suppose that $F\colon M_1 \to M_2$ is a diffeomorphism between them which is also an isometry. Then the Gaussian curvature K is the same at corresponding points.*

To see the meaning of this theorem we shall consider some examples.

(2.5) Example Let M_1 be a plane and M_2 a right circular cylinder of radius R in Euclidean space $\boldsymbol{R}^3$. If we roll the cylinder over the plane, we obtain a correspondence which does not change the length of curves or the angle between intersecting curves, hence it is an isometry. Since $K = 0$ for the plane, according to the theorem the same must be true of the cylinder. Note that they do *not* have the same second fundamental form, that is, l, m, and n do not vanish identically for the cylinder. In fact curvatures of the normal sections vary from zero to $1/R$. This depends on the imbedded shape of the surface, but K does not; it depends only on the Riemannian metric induced on M.

(2.6) Example As a second example, let M_1 be any open subset of the sphere of radius R and let M_2 be a plane. Since $K_1 \equiv 1/R^2 \neq 0$ and $K_2 \equiv 0$, the theorem implies that there exists no diffeomorphism of M_1 into M_2 that is an isometry. For example, any plane map of a portion of the globe must distort some metric properties (distance or length of curves, angles, areas, and so on). [It is interesting to note that Gauss was engaged on a surveying commission at the time he discovered his Theorema Egregium (a "most excellent theorem"). The reader is referred to the annotated translation, Gauss [1], of Gauss's famous paper for some historical comments.]

However, there do exist surfaces isometric to, but not congruent to, say, the upper hemisphere. Suppose this hemisphere to be made of a thin sheet of brass. It is intuitively clear that we may bend it by squeezing at the edge without introducing any creases (see Fig. VIII.7). This will give a surface isometric to the original since length of curves is unchanged. It follows that K is the same at corresponding points; however, the surfaces are not congruent: The circular edge has now become an oval.

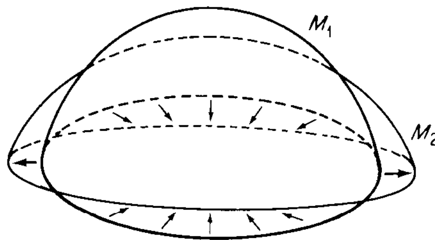


Figure VIII.7
Surface M_2 isometric to hemisphere M_1.

(2.7) Example Among the more interesting examples of (locally) isometric surfaces are the *helicoid* and the *catenoid* (Fig. VIII.8). The first surface is given parametrically by

$$(u, v) \to (u \cos v, u \sin v, v), \qquad u > 0, \quad -\infty < v < \infty.$$

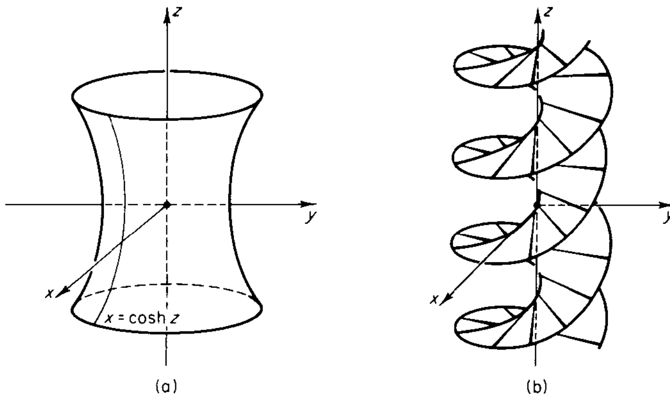


Figure VIII.8
(a) Catenoid; (b) helicoid.

It is similar in shape to a spiral staircase. On the other hand, the catenoid is obtained by revolving the catenary $x = \cosh z$ around the z-axis. We may parametrize it as

$$(z, \theta) \to (\cos \theta \cosh z, \sin \theta \cosh z, z), \qquad -\infty < z < \infty, \quad 0 < \theta < 2\pi.$$

The isometry between these surfaces is given by

$$v = \theta, \qquad u = \sinh z.$$

The verification is left as an exercise (Exercise 8).

We emphasize what by now may be obvious, namely, that Theorem 2.4 implies that the Gaussian curvature K of a two-dimensional Riemannian manifold M is determined by its structure as an *abstract* Riemannian manifold, not by its particular embedding into $\boldsymbol{R}^3$. Of course, in our presentation of K in the preceding paragraph, the Riemannian metric on the surface is given by the imbedding in $\boldsymbol{R}^3$; it is induced by the standard Riemannian metric of $\boldsymbol{R}^3$. However, according to Gauss's theorem, two very different (noncongruent) imbeddings of the same surface, say $F_1: M \to M_1 \subset \boldsymbol{R}^3$ and $F_2: M \to M_2 \subset \boldsymbol{R}^3$ have the same Gaussian curvature at each point if each imbedding induces the same Riemannian structure on M or equivalently, if $F = F_2 \circ F_1^{-1}: M_1 \to M_2$ is an isometry. This leads to the conclusion that were the theorem true, K should be computable (on the coordinate neighborhood U, φ) from the components E, F, G of the first fundamental form alone. This is the classical proof. (See, for example, Stoker, p. 139 [1].) The present proof takes advantage of the subsequent work of Riemann and uses the Riemann curvature tensor, introduced briefly in Section VII.4, together with consequences of the fundamental theorem of Riemannian geometry

both offspring of the work of Gauss and Riemann. At the same time we shall make some first steps toward investing the curvature tensor with geometric meaning.

Proof of Theorem 2.4 We remember that at a point $p \in M$ the value of the Gaussian curvature K is given by

$$K = \frac{ln - m^2}{EG - F^2},$$

where E, F, G and l, m, n are the components of the first and second fundamental forms, respectively, relative to a system of local coordinates u, v in a neighborhood U of p. The value of the ratio K is independent of the coordinates chosen although E, F, G and l, m, n are not. Let $E_1 = X_u$ and $E_2 = X_v$, where $X = X(u, v)$ gives the surface in $\mathbf{R}^3$. Then we have seen that

$$ln - m^2 = \left(\frac{\partial N}{\partial u}, E_1\right)\left(\frac{\partial N}{\partial v}, E_2\right) - \left(\frac{\partial N}{\partial u}, E_2\right)\left(\frac{\partial N}{\partial v}, E_1\right)$$

and, since $E = (E_1, E_1)$, $F = (E_1, E_2)$, and $G = (E_2, E_2)$,

$$EG - F^2 = (E_1, E_1)(E_2, E_2) - (E_1, E_2)^2.$$

Since E, F, G are the coefficients of the Riemannian metric, it is enough to show that $ln - m^2 = K(EG - F^2)$ depends only on the Riemannian metric. We shall show that

$$ln - m^2 = R(E_1, E_2, E_2, E_1),$$

where $R(X, Y, Z, W)$ is the covariant tensor of order 4 defined in Section VII.4, in which case K is given by

$$\textbf{(2.8)} \qquad K = (EG - F^2)^{-1} R(E_1, E_2, E_2, E_1) = \frac{R(E_1, E_2, E_2, E_1)}{(E_1, E_1)(E_2, E_2) - (E_1, E_2)^2}.$$

The left side is independent of local coordinates; thus, the right side is also. In fact it is easily shown that replacing E_1, E_2 at a point by any pair of vectors F_1, F_2, spanning the same plane, leaves unchanged the expression on the right-hand side of formula (2.8), which we shall prove gives K. This expression, defined at each point of an imbedded surface M, is thus independent of local coordinates on M, and moreover it depends only on the Riemannian metric. Clearly this is true of the denominator and we recall that by definition

$$(R(E_1, E_2) \cdot E_2, E_1) = (\nabla_{E_1}\nabla_{E_2}E_2 - \nabla_{E_2}\nabla_{E_1}E_2 - \nabla_{[E_1, E_2]}E_2, E_1),$$

which depends only on the Riemannian metric by the fundamental theorem of Riemannian geometry (Theorem VII.3.3). In fact in our present case, since E_1 and E_2 denote coordinate frames of local coordinates u, v, we know that $[E_1, E_2] = 0$, and so we must show only that

$$ln - m^2 = (\nabla_{E_1}\nabla_{E_2}E_2 - \nabla_{E_2}\nabla_{E_1}E_2, E_1)$$

in order to prove the theorem.

We may compute the right-hand side using the definition of $\nabla_{E_i}Z$, $i = 1, 2$ (for any tangent vector field Z), given originally in Section VII.2; namely, we take $\partial Z/\partial u$ and $\partial Z/\partial v$, project them to the tangent plane at each point of the surface, and obtain $DZ/\partial u = \nabla_{E_1}Z$ and $DZ/\partial v = \nabla_{E_2}Z$. If N denotes the unit normal, and $E_1 = X_u$ and $E_2 = X_v$, then this procedure gives

$$\nabla_{E_1}E_2 = X_{uv} - (N, X_{uv})N, \qquad \nabla_{E_2}E_2 = X_{vv} - (N, X_{vv})N.$$

Differentiating again and projecting onto the tangent plane (by subtracting the normal component of the derivative) gives

$$\nabla_{E_2}(\nabla_{E_1}E_2) = X_{vuv} - (N, X_{uv})N_v - c_1 N,$$
$$\nabla_{E_1}(\nabla_{E_2}E_2) = X_{uvv} - (N, X_{vv})N_u - c_2 N.$$

We need not compute the scalars c_1 and c_2 multiplying N since $(N, E_1) = 0$ so that these terms vanish in the final computation, in which we take an inner product of each term above with E_1. This yields for $R(E_1, E_2, E_2, E_1)$

$$(\nabla_{E_1}\nabla_{E_2}E_2 - \nabla_{E_2}\nabla_{E_1}E_2, E_1) = (X_{uvv}, X_u) - (N, X_{vv})(N_u, X_u)$$
$$- (X_{vuv}, X_u) + (N, X_{uv})(N_v, X_u).$$

This must be seen to be equal to the earlier evaluation of $ln - m^2$ above, namely,

$$ln - m^2 = (N_u, X_u)(N_v, X_v) - (N_u, X_v)(N_v, X_u).$$

Since $X_{vuv} = X_{uvv}$, applying the identities developed in the proof of Theorem 1.3, we have $(N, X_{vv}) = -(N_v, X_v)$ and $(N, X_{uv}) = -(N_u, X_v)$. This completes the proof. ∎

This proof provides an interpretation of the Riemann curvature tensor for a two-dimensional Riemannian manifold. Indeed, when F_1, F_2 are chosen at $p \in M$ so that they are *mutually perpendicular unit vectors*, then expression (2.8), which we have found for the Gauss curvature K, becomes

$$K = R(F_1, F_2, F_2, F_1).$$

Exercises

1. Prove the Lagrange identity for the inner product of the cross product of vectors of $\boldsymbol{R}^3$:
$$(X \times Y, U \times V) = (X, U)(Y, V) - (X, V)(Y, U).$$
2. Show that k_1 and k_2 are given by $H \pm (H^2 - K)^{1/2}$.
3. Show that if $K > 0$ at p, then there are no asymptotic directions, and if $K < 0$ at p, then there are two asymptotic directions and the principal directions bisect the angles made by the asymptotic directions.
4. Show that a surface M is minimal if and only if there are two asymptotic directions at each point and they are mutually orthogonal.
5. For a surface of revolution formed by revolving $z = f(x)$ around the x-axis determine when $K > 0$, when $K = 0$, and when $K < 0$. Give a sufficient condition that the surface be minimal.
6. Verify that a diffeomorphism of two Riemannian manifolds which preserves lengths of all C^1 curves is an isometry, that is, it preserves the inner product in the tangent spaces at corresponding points.
7. Verify that (2.8) is unchanged if E_1, E_2 is replaced at p by F_{1p}, F_{2p}, another basis of $T_p(M)$.
8. Verify that Example 2.7 is a local isometry as claimed.

3 Basic Properties of the Riemann Curvature Tensor

We have previously (Section VII.4) defined the curvature tensor $R(X, Y, Z, W)$ of a Riemannian manifold M. Recall that it is a covariant tensor field of order 4 whose value at any point $p \in M$ is determined as follows: Let X, Y, Z, W be vector fields whose values at p are given, say X_p, Y_p, Z_p, W_p. Then
$$R(X_p, Y_p, Z_p, W_p) = (\nabla_{X_p}\nabla_Y Z - \nabla_{Y_p}\nabla_X Z - \nabla_{[X,Y]_p}Z, W_p).$$
We have shown that this is independent of the vector fields chosen and defines a C^∞ covariant tensor field.

In the same way the vector fields X, Y define at each $p \in M$ a linear operator, the curvature operator, $R(X_p, Y_p)$ on $T_p(M)$ by the prescription
$$R(X_p, Y_p) \cdot Z_p = \nabla_{X_p}\nabla_Y Z - \nabla_{Y_p}\nabla_X Z - \nabla_{[X,Y]_p}Z,$$
which is—like the curvature tensor—linear in X, Y, Z in the sense of a $C^\infty(M)$ module, that is, if $f \in C^\infty(M)$, then
$$fR(X, Y) \cdot Z = R(fX, Y) \cdot Z = R(X, fY) \cdot Z = R(X, Y) \cdot fZ.$$
Obviously these objects are related by the equality
$$R(X, Y, Z, W) = (R(X, Y) \cdot Z, W).$$

As we now prove, the curvature tensor satisfies a number of important symmetry relations.

(3.1) Theorem *The following symmetry relations hold for the curvature tensor and curvature operator at each point, and hence for all vector fields.*

$$\text{(i)}\qquad R(X, Y)\cdot Z + R(Y, X)\cdot Z = 0,$$
$$\text{(ii)}\qquad R(X, Y)\cdot Z + R(Y, Z)\cdot X + R(Z, X)\cdot Y = 0,$$
$$\text{(iii)}\qquad (R(X, Y)\cdot Z, W) + (R(X, Y)\cdot W, Z) = 0,$$
$$\text{(iv)}\qquad (R(X, Y)\cdot Z, W) = (R(Z, W)\cdot X, Y).$$

Proof Relation (i) follows immediately from the formula above which defines the operator $R(X, Y)$. The fact that $R(X, Y, Z, W)$ is a tensor, in particular, the linearity with respect to C^∞ functions, has the following important consequence: It suffices to prove any of these statements for the vectors of a field of coordinate frames, say $E_1, \ldots, E_n$. However, for these vector fields the Lie products $[E_i, E_j] = 0$; so if X, Y, Z are chosen from among $E_1, \ldots, E_n$, then proving (ii) reduces to showing that

$$\nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) + \nabla_Y(\nabla_Z X) - \nabla_Z(\nabla_Y X) + \nabla_Z(\nabla_X Y) - \nabla_X(\nabla_Z Y) = 0.$$

By definition of Riemannian connection, $\nabla_X Y - \nabla_Y X = [X, Y] = 0$. Using this, we find that the terms on the left cancel two by two; this proves (ii). To prove (iii) we may show that the equivalent statement, $(R(X, Y)\cdot Z, Z) = 0$ for all X, Y, Z, is true. Again it is enough to do so for X, Y, Z chosen from among the vectors of the coordinate frames so that $[X, Y] = 0$. Applying the definitions, we see that

$$(R(X, Y)\cdot Z, Z) = (\nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z), Z) = 0$$

if and only if $(\nabla_X(\nabla_Y Z), Z)$ is symmetric in X, Y. Now differentiating the inner product (Z, Z) with respect to X and Y, we find that

$$Y(X(Z, Z)) = 2Y(\nabla_X Z, Z) = 2(\nabla_Y(\nabla_X Z), Z) + 2(\nabla_X Z, \nabla_Y Z),$$

from which it follows that

$$(\nabla_Y(\nabla_X Z), Z) = \tfrac{1}{2}YX(Z, Z) - (\nabla_X Z, \nabla_Y Z).$$

Since $[X, Y] = 0$, $(XY - YX)f \equiv 0$ for any function f, and in particular, taking $f = (Z, Z)$, we see that the right side is symmetric in X, Y and so also the left.

Property (iv) is derived from the first three as follows. We take the inner product of (ii) with W and obtain from this three more equations by circular permutation of X, Y, Z, W, namely,

$$\begin{aligned}
(R(X, Y)\cdot Z, W) + (R(Y, Z)\cdot X, W) + (R(Z, X)\cdot Y, W) &= 0,\\
(R(Y, Z)\cdot W, X) + (R(Z, W)\cdot Y, X) + (R(W, Y)\cdot Z, X) &= 0,\\
(R(Z, W)\cdot X, Y) + (R(W, X)\cdot Z, Y) + (R(X, Z)\cdot W, Y) &= 0,\\
(R(W, X)\cdot Y, Z) + (R(X, Y)\cdot W, Z) + (R(Y, W)\cdot X, Z) &= 0,
\end{aligned}$$

Adding the first two and subtracting the second two of these and using (i) and (iii) yields.

$$2(R(X, Y)\cdot Z, W) - 2(R(Z, W)\cdot X, Y) = 0,$$

which proves (iv). ∎

In any coordinate neighborhood U, φ we have coordinate frames $E_1, \ldots, E_n$ and we may introduce (as in Remark VII.4.5) n^4 functions of the coordinates $R_i{}^j{}_{kl}$, $1 \le i, j, k, l \le n$ by the equations

$$R(E_k, E_l)\cdot E_i = \sum_j R_i{}^j{}_{kl} E_j .$$

Similarly we may define the components R_{ijkl} of the Riemannian curvature tensor by the equations

$$R_{ijkl} = \big(R(E_k, E_l)E_i, E_j\big) = \sum_h R_i{}^h{}_{kl} g_{hj},$$

where of course $g_{ij} = (E_i, E_j)$ are the components of the Riemannian metric. By linearity both $R(X, Y)\cdot Z$ and $\big(R(X, Y)\cdot Z, W\big)$ are determined on U by these locally defined functions. The preceding theorem may be written in terms of components as follows.

(3.2) Corollary *For all* $1 \le i, j, k, l \le n$ *we have*

(i) $R_i{}^j{}_{kl} + R_i{}^j{}_{lk} = 0,$

(ii) $R_i{}^j{}_{kl} + R_k{}^j{}_{li} + R_l{}^j{}_{ik} = 0,$

(iii) $R_{ijkl} + R_{jikl} = 0,$

(iv) $R_{ijkl} = R_{klij},$

(v) $R_{ijkl} + R_{iklj} + R_{iljk} = 0.$

We remark that (v) is an immediate consequence of $R_{ijkl} = \sum_h R_i{}^h{}_{kl} g_{hj}$ and (i)–(iv).

The Riemann curvature tensor $\big(R(X, Y)\cdot Z, W\big)$ is used to define the sectional curvature, which plays an important role in the geometry of Riemannian manifolds. At any $p \in M$ we denote by π a *plane section*, that is, a two-dimensional subspace of $T_p(M)$. Such a section is determined by any pair of mutually orthogonal unit vectors X, Y at p.

(3.3) Definition The *sectional curvature* $K(\pi)$ of the section π with orthonormal basis X, Y is defined as

$$K(\pi) = -R(X, Y, X, Y) = -(R(X, Y) \cdot X, Y).$$

From the symmetry and linearity properties it is easy to see that replacing X, Y by any pair of vectors X', Y', where $X = \alpha X' + \beta Y'$ and $Y = \gamma X' + \delta Y'$ gives the relation

$$(1/\Delta^2)(R(X', Y') \cdot X', Y') = (R(X, Y) \cdot X, Y),$$

where $\Delta = \alpha\delta - \beta\gamma$, the determinant of coefficients. If X', Y' is also an orthonormal pair, then $\Delta = \pm 1$ so that the definition of $K(\pi)$ is independent of the pair used. If it is just any arbitrary linearly independent pair, then using $\Delta^2 = (X', X')(Y', Y') - (X', Y')^2$, we have

$$K(\pi) = -\frac{(R(X'Y') \cdot X', Y')}{(X', X')(Y', Y') - (X', Y')^2}. \tag{3.4}$$

In local coordinates, using $(E_i, E_j) = g_{ij}$ and the notation above,

$$K(\pi) = -\frac{\sum R_{ijkl}\alpha^i\beta^j\alpha^k\beta^l}{\sum (g_{ik}g_{jl} - g_{il}g_{jk})\alpha^i\beta^j\alpha^k\beta^l},$$

where summation is over i, j, k, l and $X' = \sum_i \alpha^i E_i$, $Y' = \sum_j \beta^j E_j$.

Although it is not obvious, the symmetry properties of the Riemann curvature tensor imply that both $(R(X, Y) \cdot Z, W)$ and $R(X, Y) \cdot Z$ are completely determined for arbitrary X, Y, Z, W if $K(\pi)$ is known for all sections π.

(3.5) Theorem *If* $\dim M \geq 3$ *and the sectional curvature is known on all sections of* $T_p(M)$, *then the Riemann curvature tensor is uniquely determined at* p.

Proof Let $R(X, Y, Z, W)$ and $\tilde{R}(X, Y, Z, W)$ be two tensors with the symmetry properties of Theorem 3.1 and let $A(X, Y, Z, W)$ be their difference. It will also be a tensor with these symmetry properties. Our assumption is that for all X, Y, $R(X, Y, X, Y) = \tilde{R}(X, Y, X, Y)$, or equivalently, $A(X, Y, X, Y) = 0$. We must show that this implies that $A(X, Y, Z, W) = 0$ for all X, Y, Z, W, that is, that $A = 0$. Let $p \in M$ and let X, Y, Z, W, etc., denote arbitrary vectors in $T_p(M)$. Then

$$0 = A(X + Y, Z, X + Y, Z) = A(X, Z, Y, Z) + A(Y, Z, X, Z),$$

so, by (3.1)(iv),

$$A(X, Z, Y, Z) = 0.$$

From this, by replacing Z by $Z + W$, we obtain

$$0 = A(X, Z + W, Y, Z + W) = A(X, Z, Y, W) + A(X, W, Y, Z)$$

to which we apply (3.1)(iv) and then (i) to obtain

$$0 = A(Y, W, X, Z) + A(X, W, Y, Z) = A(Y, W, X, Z) - A(W, X, Y, Z).$$

Renaming the variables so that Y, W, X, Z becomes X, Y, Z, W, this may be written as

$$A(X, Y, Z, W) = A(Y, Z, X, W).$$

This says that a cyclic permutation of X, Y, Z does not change the value of A. Thus (3.1)(iii) reduces to

$$3A(X, Y, Z, W) = 0,$$

which completes the proof. ∎

This theorem does much to establish the importance of the sectional curvature in the study of Riemannian geometry. We can also use sectional curvature to give a geometric interpretation of curvature in terms of the Gaussian curvature K of surfaces. However, to do this we will first need to complete our treatment of the equations of structure, which will be done in a later section.

We shall say that a Riemannian manifold M is *isotropic at a point* $p \in M$ if the curvature is the same constant K_p on every section at p and *isotropic* if it is isotropic at every point. Of course a two-dimensional Riemannian manifold is (trivially) isotropic.

(3.6) Corollary *If p is an isotropic point of M and U, φ is a coordinate neighborhood with coordinate frames $E_1, \ldots, E_n$ and Riemannian metric $g_{ij} = (E_i, E_j)$, then*

$$R_{ijkl} = -K_p(g_{ik}g_{jl} - g_{il}g_{jk}) \qquad \text{at} \quad p.$$

Proof It is easy to check that the right side defines a tensor of order 4 on $T_p(M)$ with the same symmetry properties as $R(X, Y, Z, W)$ and with constant value on all sections. The corollary then follows from the uniqueness theorem (Theorem 3.5). ∎

(3.7) Definition An isotropic Riemannian manifold is called a manifold of *constant curvature* if K_p is the same at every point.

An example is Euclidean space where $K_p \equiv 0$. This concept will be discussed more fully in a later section.

We saw in Chapter V that there exist algebraic operations on tensors on a vector space V which yield new tensors on V. Addition and multiplication of tensors as well as the operators $\mathscr{A}$ and $\mathscr{S}$ are examples. The systematic study of these operations is a branch of linear (or multilinear) algebra. It is important to differential geometry because each such operation has an immediate counterpart in tensor fields on a manifold. This is treated systematically in many of the references, at the very beginning, for example, in the books of Sternberg [1] and Kobayashi and Nomizu [1]. We will content ourselves with examples showing how the curvature tensor yields other related tensors.

Let $R(X, Y, Z, W)$ denote the curvature tensor on a Riemannian manifold M. We shall use this curvature tensor to define a (covariant) tensor field $S(X, Y)$ of order 2 and a (scalar) function on M. Let $p \in M$ and let $F_{1p}, \ldots, F_{np}$ be an orthonormal basis at p. Then it is left as an exercise to verify that

$$S_p(X_p, Y_p) = \sum_{i=1}^{n} R(F_{ip}, X_p, Y_p, F_{ip}) = \sum_{i=1}^{n} \left(R(F_{ip}, X_p) \cdot Y_p, F_{ip}\right)$$

is independent of the choice of orthonormal basis and defines a symmetric, C^∞, covariant tensor field S on M.

(3.8) Definition The tensor field $S(X, Y)$ is called the *Ricci curvature* of M. If there is a constant c such that

$$S(X, Y) = c(X, Y),$$

that is, $S(X, Y)$ is a constant multiple of the Riemannian metric on M, then M is called an *Einstein manifold*. The function r on M, defined by

$$r(p) = \sum_{i,j=1}^{n} R(F_{ip}, F_{jp}, F_{jp}, F_{ip}) = \sum_{j=1}^{n} S(F_{jp}, F_{jp}),$$

is called the *scalar curvature* of M.

Spaces of constant curvature are examples of Einstein manifolds (Exercise 6). A further example is given by the corollary to the following theorem.

(3.9) Theorem *On a compact Lie group G with a bi-invariant Riemannian metric, the sectional curvatures at e (hence everywhere) are given by the formula*

$$K(\pi_e) = -R(X_e, Y_e, X_e, Y_e) = +\tfrac{1}{4}([X, Y], [X, Y]),$$

where X, Y are an orthonormal pair of left-invariant vector fields spanning the section π_e at e. The curvature operator is similarly given at e, hence at all points by

$$R(X, Y) \cdot Z = -\tfrac{1}{4}[[X, Y], Z]$$

with X, Y, Z left-invariant vector fields.

Proof We proved in Theorem VII.8.12, that for left-invariant vector fields X, Y, the connection of a bi-invariant metric on G is given by $\nabla_X Y = \tfrac{1}{2}[X, Y]$. Applying first the definition and then the Jacobi identity, we obtain

$$\begin{aligned} R(X, Y) \cdot Z &= \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X, Y]} Z \\ &= \tfrac{1}{4}[X, [Y, Z]] - \tfrac{1}{4}[Y, [X, Z]] - \tfrac{1}{2}[[X, Y], Z] \\ &= \tfrac{1}{4}[Z, [X, Y]] = -\tfrac{1}{4}[[X, Y], Z]. \end{aligned}$$

We also know that for left-invariant vector fields U, V, W on G the following identity holds, according to Lemma VI.8.12,

$$([U, V], W) = (U, [V, W]).$$

Thus, if X, Y are left-invariant and are an orthonormal basis at e of π, a plane section, the sectional curvature is

$$K(\pi) = -R(X, Y, X, Y) = \tfrac{1}{4}([[X, Y], X], Y) = \tfrac{1}{4}([X, Y], [X, Y]). \quad \blacksquare$$

If $\mathfrak{g}$ is a Lie algebra and $X \in \mathfrak{g}$, then $\operatorname{ad}(X)$ denotes the linear mapping of $\mathfrak{g}$ defined by $\operatorname{ad} X(Y) = [X, Y]$. By Exercise VI.8.8, $\operatorname{ad} X = 0$ if and only if $X \in \mathfrak{c}$, the center of $\mathfrak{g}$. We shall say that a compact Lie group G is *semisimple* if the center of its Lie algebra is $\{0\}$ or, equivalently (Exercise 5), if the center of G is discrete.

(3.10) Corollary *Let G be as above and X, Y, Z be left-invariant vector fields. Then the Ricci tensor $S(X, Y)$ is given by the formula*

$$S(X, Y) = -\tfrac{1}{4} \operatorname{tr}(\operatorname{ad} X \circ \operatorname{ad} Y)$$

and is positive semi-definite and bi-invariant on G. Each compact semisimple G is an Einstein manifold relative to any bi-invariant Riemannian metric.

Proof Using the formula above we see that the linear operator $Z \to R(Z, Y) \cdot X$ on G is defined at e for the left-invariant vector field by

$$R(Z, Y) \cdot X = -\tfrac{1}{4}(\operatorname{ad} X)(\operatorname{ad} Y) \cdot Z.$$

According to Exercise 3, an alternative definition of $S(X, Y)$ is that it is the trace of the linear mapping $Z \to R(Z, X) \cdot Y$ on the tangent space at each

point. Then, since $S(X, Y) = S(Y, X)$, the formula of the theorem holds. On the other hand, if $F_1, \ldots, F_n$ is an orthonormal basis of left-invariant vector fields, then the formula used in the proof above

$$(\text{ad } X \cdot F_i, F_j) = ([X, F_i], F_j) = -(F_i, [X, F_j]) = -(F_i, \text{ad } X \cdot F_j)$$

shows that the matrix (a_{ij}) of ad X, relative to this basis, is skew symmetric. Hence

$$\text{tr ad } X \text{ ad } X = \sum_{i,j} a_{ij} a_{ji} = -\sum_{i,j} a_{ij}^2.$$

Therefore, $4S(X, X) = -\text{tr ad } X \text{ ad } X = \sum a_{ij}^2 \geq 0$ with equality holding only when ad $X = 0$. Hence $S(X, Y)$ is positive semidefinite. Moreover, if G is semisimple, it is positive definite. It is clearly left-invariant: when X, Y, Z are left-invariant so is $R(Z, Y) \cdot X$ and $S(X, Y)$, its trace. This means that $S(X, Y)$ is a bi-invariant Riemannian metric on a semisimple G. However two bi-invariant metrics can differ only by a scalar multiple. It follows that with a bi-invariant metric, G is Einstein. Note that this corresponds, except for a constant factor, to the metric on $SO(n)$ of Example VII.8.6. ∎

Exercises

1. Prove that expression (3.4) depends only on the plane π determined by the vectors X', Y'.
2. Show that for any orthonormal basis $F_1, \ldots, F_n$ at $p \in M$, the values of $S(X_p, Y_p) = \sum_{i=1}^n R(F_{ip}, X_{ip}, Y_p, F_{ip})$ is independent of the choice of the orthonormal basis and that this formula defines a C^∞-tensor field $S(X, Y)$ as claimed in Definition 3.8. Verify that $S(X, Y) = S(Y, X)$.
3. Show that $S(X_p, Y_p)$ is the trace of the linear operator taking $Z_p \in T_p(M)$ to $R(Z_p, X_p) \cdot Y_p \in R_p(M)$ and use this to show that $S(X, Y)$ is C^∞ for all C^∞-vector fields X, Y.
4. Show that on a compact Lie group G with a bi-invariant Riemannian metric, the curvature is identically zero if and only if G is Abelian.
5. Show that if $\mathfrak{g}$ is the Lie algebra of a Lie group G (compact or not), then the center of G is discrete if and only if the center of $\mathfrak{g}$, $\mathfrak{c} = \{X \in \mathfrak{g} \mid [X, Y] = 0 \ \forall Y \in \mathfrak{g}\}$, is $\{0\}$.
6. Show that a Riemannian manifold of constant curvature is Einstein.

4 The Curvature Forms and the Equations of Structure

We now return to the viewpoint of Section VII.4. Let U be a neighborhood on the Riemannian manifold M such that on U is defined a C^∞ family of coframes $\theta^1, \ldots, \theta^n$ and thus, automatically, a dual C^∞ family of frames $E_1, \ldots, E_n$. They may or may not be coordinate frames of a coordinate neighborhood U, φ. The components of the Riemann metric on U are still

denoted by $g_{ij} = (E_i, E_j)$ however, and according to Theorem VII.4.6, there exist *uniquely* determined one-forms θ_i^j on U satisfying

(i) $d\theta^i = \sum_j \theta^j \wedge \theta_j^i, \qquad 1 \le i \le n,$
(ii) $dg_{ij} = \sum_k \theta_i^k g_{kj} + \sum_k g_{ik}\theta_j^k, \qquad 1 \le i, j \le n.$

[We remark that by defining $\theta_{ij} = \sum_k \theta_i^k g_{kj}$, the equations (ii) assume the simpler form $dg_{ij} = \theta_{ij} + \theta_{ji}$.] In the special case where the frames are orthonormal, that is, $g_{ij} = \delta_{ij}$, we will use ω^i, ω_i^j instead of $\theta^{i, \theta_{ij}}$. Then (ii) becomes $0 = \omega_i^j + \omega_j^i$, $1 \le i, j \le n$.

The forms θ_i^j determine, and are determined by the Riemannian connection. Thus if $\nabla_{E_i} E_j = \sum_k \Gamma_{ij}^k E_k$, then $\theta_i^j = \sum_k \Gamma_{ki}^j \theta^k$ equivalently, $\nabla_X E_j = \sum_k \theta_j^k(X) E_k$. The one-forms θ_j^k, $1 \le j, k \le n$, are called the *connection forms.* (A word of caution: $\Gamma_{ij}^k = \Gamma_{ji}^k$ only if $E_1, \ldots, E_n$ satisfy $[E_i, E_j] = 0$, as is the case for *coordinate* frames. This symmetry was derived from $\nabla_{E_i} E_j - \nabla_{E_j} E_i = [E_i, E_j]$, which we have made part of the definition of Riemannian connection; it is equivalent to (i) above.)

Now suppose that $R_i{}^j{}_{kl}$, $1 \le i, j, k, l \le n$ are the components of the curvature (as an endomorphism) relative to the given frames, that is, $R(E_k, E_l) \cdot E_i = \sum_j R_i{}^j{}_{kl} E_j$. Then we define n^2 two-forms Ω_i^j, $1 \le i, j \le n$ by

$$\Omega_i^j = \sum_{1 \le k < l \le n} R_i{}^j{}_{kl} \theta^k \wedge \theta^l = \tfrac{1}{2} \sum_{k, l = 1}^{n} R_i{}^j{}_{kl} \theta^k \wedge \theta^l.$$

It follows that

$$\sum_{j=1}^{n} \Omega_i^j(E_k, E_l) E_j = \sum_{j=1}^{n} R_i{}^j{}_{kl} E_j = R(E_k, E_l) \cdot E_i$$

and by linearity this extends to any vector fields X, Y so that

$$R(X, Y) \cdot E_i = \sum_j \Omega_i^j(X, Y) E_j;$$

thus $(\Omega_i^j(X, Y))$ is the matrix of the curvature operator relative to the basis $E_1, \ldots, E_n$. Note that the properties of $R(X, Y) \cdot Z$ imply that $\Omega_i^j(X, Y)$ at p depend only on the values of X and Y at p, not on the vector fields; obviously $\Omega_i^j(X, Y) = -\Omega_i^j(Y, X)$. These n^2 forms Ω_i^j on U_j are called the *curvature forms*; they depend on the Riemannian metric and on the particular frame-field we use on U. The following result shows the relation between these forms and the connection forms.

(4.1) Theorem *Using the notation above, the forms Ω_i^j on U are defined by the equations*

(4.2) $\qquad \Omega_i^j = d\theta_i^j - \sum_{k=1}^{n} \theta_i^k \wedge \theta_k^j, \qquad 1 \le i, j \le n.$

Proof It is sufficient to verify that on any vector fields X, Y on U the value of the two-forms on each side of the equation is the same. This is equivalent to showing that

$$R(X, Y) \cdot E_i = \sum_j \left(\left(d\theta_i^j - \sum_k \theta_i^k \wedge \theta_k^j\right)(X, Y)\right)E_j, \qquad i = 1, \ldots, n.$$

By definition,

$$R(X, Y) \cdot E_i = \nabla_X(\nabla_Y E_i) - \nabla_Y(\nabla_X E_i) - \nabla_{[X, Y]} E_i,$$

which may be rewritten

$$R(X, Y) \cdot E_i = \nabla_X\left(\sum_j \theta_i^j(Y)E_j\right) - \nabla_Y\left(\sum_j \theta_i^j(X)E_j\right) - \sum_j \theta_i^j([X, Y])E_j.$$

Since $\theta_i^j(Y)$ and $\theta_i^j(X)$ are functions, the right-hand side is equal to

$$\sum_j (X(\theta_i^j(Y)) - Y(\theta_i^j(X)) - \theta_i^j([X, Y]))E_j$$
$$+ \sum_{j,k} \theta_i^j(Y)\theta_j^k(X)E_k - \sum_{j,k} \theta_i^j(X)\theta_j^k(Y)E_k.$$

Applying Lemma V.8.4, this becomes

$$\sum_j \left\{d\theta_i^j(X, Y) - \sum_k [\theta_i^k(X)\theta_k^j(Y) - \theta_i^k(Y)\theta_k^j(X)]\right\}E_j,$$

which shows that, as claimed,

$$R(X, Y) \cdot E_i = \sum_j \left(d\theta_i^j - \sum_k \theta_i^k \wedge \theta_k^j\right)(X, Y) \ E_j.$$

This completes the proof. ∎

(4.3) Remark In summary, we have the following facts. Let U be any open subset of a Riemannian manifold M on which is defined a field of coframes $\theta^1, \ldots, \theta^n$. Let $E_1, \ldots, E_n$ denote the uniquely determined dual frame-field and let $g_{ij} = (E_i, E_j)$ on U. Then there exist n^2 uniquely determined one-forms θ_i^j on U satisfying conditions (i) and (ii) of the first paragraph of this section. They determine the two-forms Ω_i^j, and hence the curvature on U, by (4.2). Equations (i), (ii), and (4.2) are known as the *equations of structure*; they are due to Elie Cartan, who made extensive use of them. As noted above, it is often convenient to write $\theta_{ij} = \sum_s \theta_i^s g_{sj}$ so that (ii) takes a simpler form. We may define, similarly, $\Omega_{ij} = \sum_s \Omega_i^s g_{sj}$; then $\Omega_{ij} = \frac{1}{2}\sum_{k,l} R_{ijkl}\theta^k \wedge \theta^l$ since we have previously seen that $R_{ijkl} = \sum_s g_{js} R_i{}^s{}_{kl}$, where $R_{ijkl} = R(E_k, E_l, E_i, E_j)$ by Definition VII.4.5. The symmetry properties of Corollary 3.2 imply that $\Omega_{ij} = -\Omega_{ji}$.

In the event that the frame-field is orthonormal, that is, consists of vectors $E_1, \ldots, E_n$ with $(E_i, E_j) = \delta_{ij}$, then as noted above, (i) and (ii) simplify; moreover, $\Omega_{ij} = \Omega_i^j$, $R_{ijkl} = R_i{}^j{}_{kl}$ and $\omega_i^j = \omega_{ij}$. Recapitulating the remarks above we have the following corollary.

(4.4) Corollary *The forms $\omega^1, \ldots, \omega^n$ dual to a field of orthonormal frames determine uniquely a set of one-forms ω_i^j, $1 \le i, j \le n$, satisfying*

(i) $d\omega^i = \sum_k \omega^k \wedge \omega_k^i$

and

(ii) $\omega_i^j + \omega_j^i = 0$

and we have

(iii) $d\omega_i^j - \sum_k \omega_i^k \wedge \omega_k^j = \sum_{k<l} R_i{}^j{}_{kl}\omega^k \wedge \omega^l = \Omega_i^j = \Omega_{ij}$.

Relative to these frames the matrix $(\Omega_{ij}(X, Y))$ of the curvature operator $R(X, Y)$ is a skew-symmetric matrix.

(4.5) Corollary *Let Γ_{ij}^k denote the coefficients of the connection forms relative to coordinate frames $E_1, \ldots, E_n$ of a coordinate neighborhood U, φ, that is, $\theta_j^k = \sum_l \Gamma_{lj}^k \theta^l$ with $\theta^1, \ldots, \theta^n$ being dual to $E_1, \ldots, E_n$. Then $\Gamma_{ij}^k = \Gamma_{ji}^k$ and*

$$R_i{}^j{}_{kl} = \frac{\partial \Gamma_{il}^j}{\partial x^k} - \frac{\partial \Gamma_{ik}^j}{\partial x^l} - \sum_h (\Gamma_{ik}^h \Gamma_{hl}^j - \Gamma_{il}^h \Gamma_{hk}^j).$$

Proof According to the theorem

$$\Omega_i^j = d\theta_i^j - \sum_h \theta_i^h \wedge \theta_h^j;$$

hence

$$\Omega_i^j = \sum_l (d\Gamma_{li}^j \wedge \theta^l + \Gamma_{li}^j \, d\theta^l) - \sum_{k,l} \sum_h \Gamma_{ki}^h \Gamma_{lh}^j \theta^k \wedge \theta^l.$$

Now $\Gamma_{ij}^k = \Gamma_{ji}^k$ since $[E_i, E_j] = 0$ for coordinate frames, and it follows that $d\theta^l = \sum_j \theta^j \wedge \theta_j^l = \sum_{i,j} \Gamma_{ij}^l \theta^j \wedge \theta^i = 0$ (since $\theta^j \wedge \theta^i = -\theta^i \wedge \theta^j$). Therefore the second equation above may be written as

$$\frac{1}{2} \sum_{k,l=1}^n R_i{}^j{}_{kl} \theta^k \wedge \theta^l = \sum_{k,l} \frac{1}{2}\left(\frac{\partial \Gamma_{li}^j}{\partial x^k} - \frac{\partial \Gamma_{ki}^j}{\partial x^l}\right)\theta^k \wedge \theta^l$$

$$- \frac{1}{2} \sum_{k,l} \sum_h (\Gamma_{ki}^h \Gamma_{lh}^j - \Gamma_{li}^h \Gamma_{kh}^j)\theta^k \wedge \theta^l.$$

Since we have made the coefficients on both left and right skew-symmetric in the indices k, l, these equations imply equality of coefficients. We use the symmetry of Γ_{ij}^k in i, j, the fact that $\theta^k \wedge \theta^l = -\theta^l \wedge \theta^k$, and change of index of summation where necessary to obtain the (standard) formula of the corollary. ∎

Consider the special case in which dim $M = 2$ and assume, moreover, that only orthonormal frames are used.

(4.6) Corollary *If* dim $M = 2$, *then* $d\omega_1^2 = \Omega_1^2 = -K\omega^1 \wedge \omega^2$, *where* K *is the Gaussian curvature of* M.

Proof In proving Gauss's Theorema Egregium we saw that if E_1, E_2 are orthonormal unit vectors, then

$$K = -R(E_1, E_2, E_1, E_2) = -(R(E_1, E_2) \cdot E_1, E_2) = -R_{1212}.$$

On the other hand, since $g_{ij} = (E_i, E_j) = \delta_{ij}$ we have

$$\Omega_1^2 = \Omega_{12} = +R_{1212}\,\omega^1 \wedge \omega^2.$$

Since $\omega_i^j + \omega_j^i = 0$, $\omega_1^1 = 0 = \omega_2^2$. Thus $\sum_{k=1}^2 \omega_1^k \wedge \omega_k^2 = 0$ and $d\omega_1^2 = \Omega_1^2$ by Corollary 4.4. This completes the proof. ∎

Note that these equations are independent of the *particular* orthonormal frame field on $U \subset M$. We shall now use this to give a geometric interpretation of sectional curvature. Let π be a plane section at a point p of M, a Riemannian manifold, and let N_p be an open, two-dimensional submanifold of M consisting of geodesic arcs through p and tangent at p to the section π.

(4.7) Theorem *If we use on* N_p *the Riemannian metric induced by that of* M, *then the sectional curvature* $K(\pi)$ *is equal to the Gaussian curvature of* N_p *at* p.

Proof Let $U = \exp_p B_\varepsilon$ be a normal neighborhood of p, that is, we choose $\varepsilon > 0$ such that $B_\varepsilon = \{X_p \in T_p(M) \mid \|X_p\| < \varepsilon\}$ is mapped diffeomorphically onto an open set $U \subset M$. The plane section π corresponds to a two-dimensional subspace $V_\pi \subset T_p(M)$ and we may suppose that N_p is the image of $V_\pi \cap B_\varepsilon$. Since U is a normal neighborhood, it is covered simply by the geodesics of length ε issuing from p; they are given by $\exp_p tX_p$, $0 \le t \le \varepsilon$, for each X_p with $\|X_p\| = 1$. Now choose an orthonormal basis $E_{1p}, \ldots, E_{np}$ of $T_p(M)$, with E_{1p}, E_{2p} a basis of V_π. Then $(x^1, \ldots, x^n) \to \exp_p(\sum x^i E_{ip})$ establishes a system of normal coordinates on U, the coordinate map φ being the inverse of the above. Thus N_p is described by $x^3 = \cdots = x^n = 0$, and $U \cap N_p$, φ is a coordinate system on N_p with x^1, x^2 as coordinates. Let $E_1, \ldots, E_n$ denote the *coordinate* frames; they agree at p with the given frame and E_1, E_2 are tangent to N_p everywhere on N_p. We denote the dual coframes by $\theta^1, \ldots, \theta^n$, with $\theta_j^k = \sum_i \Gamma_{ij}^k \theta^i$ as connection forms. Note that $\Gamma_{ij}^k(0) = 0$, that is, $\theta_j^k = 0$ at $p \in U$. This was proved in Remark VII.6.8.

From those frames, by the Gram–Schmidt process† we obtain a family of orthonormal frames $F_1, \ldots, F_n$ in U with the property that F_1, F_2 are a

† See, e.g., Hoffman and Kunze [1].

linear combination of E_1, E_2 and thus tangent to N_p at each of its points. We denote by $\omega^1, \ldots, \omega^n$ the dual coframes to $F_1, \ldots, F_n$ and by ω_i^j the corresponding connection forms; they satisfy the equations $\omega_i^j + \omega_j^i = 0$ and $d\omega^i = \sum_k \omega^k \wedge \omega_k^i$. We shall see that for $j > 2$, $\omega_1^j = \omega_2^j = 0$ at p. First recall that at p, $\nabla_{X_p} E_i = \sum_j \theta_i^j(X_p)E_j = 0$ and $\nabla_{X_p} F_i = \sum \omega_i^j(X_p)F_j$. Now for $i = 1, 2$, $F_i = a_i^1 E_1 + a_i^2 E_2$ and so

$$\nabla_{X_p} F_i = (X_p a_i^1)E_1 + (X_p a_i^2)E_2 + a_i^1 \nabla_{X_p} E_1 + a_i^2 \nabla_{X_p} E_2 .$$

Since $\Gamma_{ij}^k(0) = 0$, the last two terms vanish so that for $i = 1, 2$, $\nabla_{X_p} F_i$ is a linear combination of E_1 and E_2, and hence of F_1 and F_2. Thus $\nabla_{X_p} F_i = \omega_i^1(X_p)F_1 + \omega_i^2(X_p)F_2$ for $i = 1, 2$ and $\omega_i^j(X_p) = 0$ for $i = 1, 2$ and $j > 2$ as claimed.

Now we denote by $I: N_p \to M$ the imbedding and let $\tilde{\omega}^i = I^*\omega^i$, $\tilde{\omega}_i^j = I^*\omega_i^j$. Since I^* is a homomorphism of $\bigwedge(M) \to \bigwedge(N_p)$ and commutes with d, we know that $d\tilde{\omega}^i = \sum_k \tilde{\omega}^k \wedge \tilde{\omega}_k^i$ and $\tilde{\omega}_i^j + \tilde{\omega}_j^i = 0$. Moreover, $\tilde{\omega}^i = 0$ for $i > 2$, since F_1, F_2 spans the tangent space to N_p and $\tilde{\omega}^i(F_j) = (I^*\omega^i)(F_j) = \omega^i(I_* F_j) = \omega^i(F_j) = 0$ if $j = 1$ or $j = 2$ and $i > j$. Thus $\tilde{\omega}^1$, $\tilde{\omega}^2$ are dual to F_1, F_2 restricted to N_p and together with $\tilde{\omega}_1^2 = -\tilde{\omega}_2^1$ satisfy equations (i) and (ii) (of Remark 4.3), which determine the connection forms uniquely. It follows from Corollary 4.6 that $d\tilde{\omega}_1^2 = -K\tilde{\omega}^1 \wedge \tilde{\omega}^2$. On the other hand, we have on M

$$d\omega_1^2 = \sum_k \omega_1^k \wedge \omega_k^2 + \sum_{k<l} R_{12kl}\,\omega^k \wedge \omega^l$$

and applying I^* to both sides and evaluating at p yields the equality (at p):

$$d\tilde{\omega}_1^2 = R_{1212}\,\tilde{\omega}^1 \wedge \tilde{\omega}^2.$$

It follows that the sectional curvature $\hat{K}(\pi) = R_{1212} = K_p$, the Gaussian curvature at p of the surface N_p. This completes the proof. ∎

(4.8) Corollary *Let M be an n-sphere of radius a in $\mathbf{R}^{n+1}$ with the Riemannian metric induced from $\mathbf{R}^{n+1}$. Then M has constant sectional curvature $1/a^2$.*

Proof If p is a point of M, then the geodesics through p tangent to a plane π in $T_p(M)$ are great circles and form a 2-sphere of radius a. We have seen that the Gaussian curvature of such a 2-sphere is $1/a^2$ so the corollary follows from the theorem. ∎

We have made a distinction between isotropic manifolds and manifolds of constant curvature. A theorem of Schur [1] shows that this distinction is artificial.

(4.9) Theorem *If M is a connected, isotropic Riemannian manifold and* $\dim M \geq 3$, *then M has constant curvature.*

Proof If we let K_p be the value of the sectional curvature at p—the same on all sections by hypothesis—then we must show that this function on M is constant, that is, $dK = 0$. Let U be a neighborhood of $p \in M$ with an orthonormal frame field and let $\omega^1, \ldots, \omega^n$ be the dual coframe field. Using the expression for R_{ijkl} in Corollary 3.6, which now becomes $R_{ijkl} = -K(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})$, we have $\Omega_i^j = \Omega_{ij} = -K\omega^i \wedge \omega^j$, in which K depends only on p, not on the (orthonormal) frames used. Taking the exterior derivative of the structure equation

$$d\omega_i^j = \sum \omega_i^k \wedge \omega_k^j + \Omega_i^j,$$

we obtain

$$0 = \sum (d\omega_i^k \wedge \omega_k^j - \omega_i^k \wedge d\omega_k^j) - dK \wedge \omega^i \wedge \omega^j - K\, d\omega^i \wedge \omega^j + K\omega^i \wedge d\omega^j.$$

Substituting for $d\omega_i^k$, $d\omega^i$, and so on, from Corollary 4.4 and simplifying gives

$$dK \wedge \omega^i \wedge \omega^j = 0,$$

which holds for every $i, j = 1, \ldots, n$. Since $dK = K_1\omega^1 + \cdots + K_n\omega^n$, a linear combination of $\omega^1, \ldots, \omega^n$, and since $\omega^l \wedge \omega^i \wedge \omega^j \neq 0$ if l, i, j are distinct, this can only hold if $dK = 0$ on U, a neighborhood of p. Because p is arbitrary, $dK = 0$ and K is constant. ∎

According to Corollary 4.8, the sphere of radius a with the Riemannian metric induced by the Euclidean space with contains it has constant positive curvature. Euclidean space itself with its standard Riemannian metric has curvature identically zero, since with the usual coordinates $\Gamma_{ij}^k \equiv 0$ and $R_{ijkl} \equiv 0$. It remains to give an example of a manifold of constant negative curvature of arbitrary dimension. This will be done in Section 6, which is devoted to spaces of constant curvature. In the meantime, the reader should try Exercise 1.

Exercises

1. Show that the two-dimensional manifold $M = \{(x, y) \mid y > 0\}$, the upper half-plane, with Riemannian metric in the xy-coordinates which cover M given by $g_{11} = 1/y^2 = g_{22}$ and $g_{12} = 0 = g_{21}$, has Gaussian curvature $K = -1$.
2. Let $\theta^1, \ldots, \theta^n$ be n linearly independent one-forms defined on an open subset $U \subset \boldsymbol{R}^n$ and let $(g_{ij}(x))$ be a symmetric positive definite matrix whose entries are C^∞ functions on U. Show by direct computation that there exist uniquely determined functions Γ_{ijk}, $1 \leq i, j, k \leq n$ such that the n^2 forms $\theta_{jk} = \sum_{i=1}^n \Gamma_{ijk}\theta^i$ satisfy the two systems of equations:

   (i) $d\theta^k = \sum_{j,l} \theta^j \wedge \theta_{jl} g^{kl}$, and
   (ii) $dg_{ij} = \theta_{ij} + \theta_{ji}$.

5 Differentiation of Covariant Tensor Fields

Until this point we have used covariant differentiation D/dt on Riemannian manifolds and the associated Riemannian connection ∇ only to differentiate vector fields—either along curves or in various directions X_p at a point p of the manifold M. However, once this has been done, it is a relatively simple matter to extend the procedure to tensor fields; for the basic difficulty—lack of a method to compare vectors, tensors, and so on, on tangent spaces at nearby points q to a given $p \in M$—has been surmounted. In Euclidean space we compare $T_q(M)$ and $T_p(M)$ by parallel translation; we have also used the local one-parameter group of transformations generated by a vector *field* X to define the Lie derivative L_X. Neither procedure is available in the general Riemannian case; but there we do have parallel transport along a curve from p to q and that is what we now apply. For convenience, and since it is all that we need, we restrict our consideration to covariant tensor fields. Just as in our earlier treatment of differentiation of vector fields on a Riemannian manifold, we first differentiate a covariant tensor field along a curve and then, later, determine its derivative in various directions X_p at a point p of the manifold.

We consider, then, a covariant tensor field Φ of order r on the Riemannian manifold M, $\Phi \in \mathscr{T}^r(M)$, and we suppose given a curve $p(t)$, $a \le t \le b$, on M of differentiability class C^1 at least. Let $\Phi_{p(t)}$ denote the restriction of Φ to $p(t)$. Then $\Phi_{p(t)} \in \mathscr{T}^r(T_{p(t)}(M))$, that is, $\Phi_{p(t)}$ is a tensor field *along* $p(t)$. Using Theorem VII.3.12 and Remark VII.3.13, we denote by τ_t parallel translation along $p(t)$ from a fixed point $p(t_0)$ of the curve:

$$\tau_t \colon T_{p(t_0)}(M) \to T_{p(t)}(M).$$

This is an isomorphism of these tangent spaces and is uniquely determined by the curve $p(t)$ and the Riemannian structure. It is exactly what is needed to define the derivative of Φ at $p(t_0)$.

(5.1) Definition With the preceding notation, the derivative $D\Phi/dt$ of the tensor Φ along the curve is defined at the point $p(t_0)$ by

$$\left(\frac{D\Phi}{dt}\right)_{t_0} = \lim_{t \to t_0} \frac{1}{t - t_0} \left(\tau_t^* \Phi_{p(t)} - \Phi_{p(t_0)}\right).$$

As thus defined $(D\Phi/dt)_{t_0}$ is a covariant tensor of order r on the vector space $T_{p(t_0)}(M)$. In fact, given any set of r vectors $X^1_{p(t_0)}, \ldots, X^r_{p(t_0)} \in T_{p(t_0)}(M)$, then $D\Phi/dt$ at $p(t_0)$ is the limit as $t \to t_0$ of the expression

$$\frac{1}{t - t_0} \left(\tau_t^* \Phi_{p(t)}(X^1_{p(t_0)}, \ldots, X^r_{p(t_0)}) - \Phi_{p(t_0)}(X^1_{p(t_0)}, \ldots, X^r_{p(t_0)})\right)$$

which for each value of t near t_0 is a multiple [by $1/(t - t_0)$] of the difference of two tensors $\tau_t^*\Phi_{p(t)}$ and $\Phi_{p(t_0)}$ on $T_{p(t_0)}(M)$. Since *both* are covariant r tensors on the same vector space, it follows that the limit is also such a tensor. Repeating this procedure at each t_0 on the interval (a, b) gives a covariant tensor field $D\Phi/dt$ along $p(t)$, provided that suitable differentiability conditions are satisfied. We mean by this that for any C^k family of vector fields $X_t^i = X_{p(t)}^i$, $i = 1, \ldots, r$, defined along the C^k curve $p(t)$, the value of $D\Phi/dt$ on them,

$$\frac{D\Phi}{dt}(X_t^1, \ldots, X_t^r), \qquad a < t < b,$$

should be a function of class C^{k-1} (C^∞ when $k = \infty$) of the variable t. In particular, this should be true in the most frequent situation: $X^1, \ldots, X^r$ are C^∞-vector fields on M and $X_t^1, \ldots, X_t^r$ are their restrictions to the curve $p(t)$. In order to see that this is indeed a consequence of our definition and to derive computational formulas, we prove the following lemma. For convenience—and since it is the most important case—we suppose Φ is C^∞.

(5.2) Lemma *Let Φ be a C^∞-covariant tensor field of order r on M and let $p(t)$, $a < t < b$, be a curve of class C^k, $k \geq 1$, on M. If $X_t^1, \ldots, X_t^r \in T_{p(t)}(M)$ are vector fields of class C^k along the curve, then for each t_0 on the interval (a, b) we have*

$$\text{(5.3)} \qquad \left(\frac{D\Phi}{dt}\right)_{t_0}(X_{t_0}^1, \ldots, X_{t_0}^r) = \left(\frac{d}{dt}[\Phi_{p(t)}(X_t^1, \ldots, X_t^r)]\right)_{t=t_0} - \sum_{i=1}^{r} \Phi_{p(t_0)}\left(X_{t_0}^1, \ldots, \left(\frac{DX^i}{dt}\right)_{t_0}, \ldots, X_{t_0}^r\right).$$

Proof Before beginning the proof we note that it will indeed establish the fact that $D\Phi/dt$ evaluated on C^k-vector fields along the curve is differentiable of class C^{k-1} at least. If $k = \infty$, as will often be the case, then $D\Phi/dt$ will be a C^∞-tensor field along the curve, that is, its value on C^∞-vector fields will be a C^∞ function of t. Although Φ itself has been assumed C^∞, in fact it is sometimes convenient to consider curves of lower differentiability class, which in turn lowers the class of $D\Phi/dt$. The assertion of the lemma is easy to prove. By definition we have

$$\left(\frac{D\Phi}{dt}\right)_{t_0} = \lim_{t \to t_0} \frac{1}{t - t_0}(\tau_t^*\Phi_{p(t)}(X_{t_0}^1, \ldots, X_{t_0}^r) - \Phi_{p(t_0)}(X_{t_0}^1, \ldots, X_{t_0}^r))$$

$$= \lim_{t \to t_0} \frac{1}{t - t_0}(\Phi_{p(t)}(\tau_t(X_{t_0}^1), \ldots, \tau_t(X_{t_0}^r)) - \Phi_{p(t_0)}(X_{t_0}^1, \ldots, X_{t_0}^r)).$$

Then for each $i = 1, \ldots, r$ in turn we subtract and add

$$\Phi_{p(t)}(X_t^1, X_t^2, \ldots, X_t^i, \tau_t(X_{t0}^{i+1}), \ldots, \tau_t(X_{t0}^r)).$$

Rearranging and collecting terms and using both linearity at $p(t)$ and the continuity of the tensor Φ, we may rewrite the defining equation

$$\left(\frac{D\Phi}{dt}\right)_{t_0} = \sum_{i=1}^{r} \Phi_{p(t)}\left(X_t^1, \ldots, \lim_{t \to t_0} \frac{1}{t - t_0} (\tau_t(X_{t0}^i) - X_t^i), \tau_t(X_{t0}^{i+1}), \ldots, \tau_t(X_{t0}^r)\right)$$

$$+ \lim_{t \to t_0} \frac{1}{t - t_0} (\Phi_{p(t)}(X_t^1, \ldots, X_t^r) - \Phi_{p(t_0)}(X_{t0}^1, \ldots, X_{t0}^r)).$$

We now use the fact that for any C^k-vector field X_t along $p(t)$,

$$\lim_{t \to t_0} \frac{\tau_t(X_{t0}) - X_t}{t - t_0} = -\lim_{t \to t_0} \tau_t\left(\frac{\tau_{-t}(X_t) - X_{t0}}{t - t_0}\right) = -\tau_0\left(\frac{DX_t}{dt}\right)_{t0} = -\left(\frac{DX_t}{dt}\right)_{t0}.$$

Therefore passing to the limit in the expression for $(D\Phi/dt)_{t_0}$ completes the proof of formula (5.3) of the lemma. ∎

We can verify from the formula itself that $(D\Phi/dt)_{t_0}$ depends $\boldsymbol{R}$-linearly on the values of the vector fields $X_t^1, \ldots, X_t^r$ at $p(t_0)$ so that the formula does define an $\boldsymbol{R}$-linear function, that is, a covariant tensor of order r on the vector space $T_{p(t_0)}(M)$. This is made even clearer, however, by the following corollary—which uses the notation above.

(5.4) Corollary *Let $X_0^1, \ldots, X_0^r \in T_{p(t_0)}(M)$ be given and suppose that $X_t^1, \ldots, X_t^r$ are the uniquely determined parallel vector fields along $p(t)$, $a < t < b$, which take these values at $p(t_0)$. Then formula* (5.3) *becomes*

$$\left(\frac{D\Phi}{dt}\right)_{t_0} (X_{t_0}^1, \ldots, X_{t_0}^r) = \left(\frac{d}{dt} \Phi_{p(t)}(X_t^1, \ldots, X_t^r)\right)_{t=t_0}.$$

Proof This follows from (5.3) since by definition of X_t^i we have $DX_t^i/dt \equiv 0$, $i = 1, \ldots, r$. ∎

This corollary makes it clear that $(D\Phi/dt)_{t_0}$ depends only on the tensor field Φ and on the curve $p(t)$, $a < t < b$. In fact, it is easy to verify from the formulas above that it depends on the tangent vector $Y_{p(t_0)} = \dot{p}(t_0)$ to the curve, but not on the curve itself, more precisely, two curves through $p(t_0)$ with the same tangent vector at that point will define the same element $(D\Phi/dt)_{t_0}$ of $\mathscr{T}^r(T_{p(t_0)}(M))$. We shall state this in the form of a lemma.

(5.5) Lemma *Let Φ be as above and let $p \in M$. We suppose that $X^1, \ldots, X^r$ are C^∞-vector fields on some neighborhood U of p and let $X_p^1, \ldots, X_p^r$ denote their value at p. Let $F(t)$, $-\varepsilon < t < \varepsilon$, and $G(s)$, $-\delta < s < \delta$, be two C^1 curves*

on M such that $F(0) = p = G(0)$ and $\dot{F}(0) = Y_p = \dot{G}(0)$ is their common tangent vector at p. Then

$$\left(\frac{D\Phi}{dt}\right)_0 (X_p^1, \ldots, X_p^r) = \left(\frac{D\Phi}{ds}\right)_0 (X_p^1, \ldots, X_p^r),$$

that is, the two tensors on $T_p(M)$ defined by differentiating Φ along each of the curves is the same.

Proof Suppose that f is a C^∞ function on U. Then $f(F(t))$ is its restriction to the curve $F(t)$ and

$$\left(\frac{d}{dt} f(F(t))\right)_{t=0} = F_*\left(\frac{d}{dt}\right) f = Y_p f,$$

Similarly restricting f to $G(s)$, differentiating with respect to s, and evaluating at $s = 0$ gives $Y_p f$. Applying this to the function

$$f(q) = \Phi_q(X_q^1, \ldots, X_q^r),$$

we see that in formula (5.3) the first term in case of either curve (and derivative of Φ) is the same, namely $Y_p(\Phi(X^1, \ldots, X^r))$. On the other hand, by our original definition of $\nabla_{Y_p} X$ for a vector field X in Chapter VII we have $\nabla_{Y_p} X = (D/dt(X^i_{p(t)}))_0 = (D/ds(X^i_{p(s)}))_0$; hence the other terms in formula (5.3) agree also, which establishes the lemma. ∎

We denote the covariant tensor of order r on $T_p(M)$ which we have thus defined from differentiation of Φ along curves through p with Y_p as tangent at p by $\nabla_{Y_p}\Phi$.

(5.6) Definition The covariant r tensor on $T_p(M)$ just defined $\nabla_{Y_p}\Phi \in \mathscr{T}^r(T_p(M))$, is called the *covariant derivative* of Φ at p in the direction Y_p.

According to the facts in the proof above, the covariant derivative is given by the formula

(5.7)
$$\nabla_{Y_p}\Phi(X^1, \ldots, X^r) = Y_p(\Phi(X^1, \ldots, X^r)) - \sum_{i=1}^{r} \Phi_p(X_p^1, \ldots, \nabla_{Y_p} X^i, \ldots, X_p^r),$$

where $X^1, \ldots, X^r$ are vector fields on a neighborhood of p. Only their values at p affect the value of $\nabla_{Y_p}\Phi$ on $T_p(M)$.

(5.8) Theorem *Given $\Phi \in \mathscr{T}^r(M)$ as above, then we may define on M a C^∞-covariant tensor field Ψ of order $r + 1$ by the formula*

$$\Psi_p(X_p^1, \ldots, X_p^r; Y_p) = (\nabla_{Y_p}\Phi)(X_p^1, \ldots X_p^r).$$

Proof In view of all that has been shown above it is only necessary to prove two more facts: first that for each $p \in M$, Ψ_p is linear in the last variable —with the others fixed. Second, that for any C^∞-vector fields $X^1, \dots, X^r$, Y, the formula above defines a C^∞ function of p.

The first fact is a consequence of the linearity in Y_p of each term of (5.7) as a real-valued function on $T_p(M)$. Thus, if we fix the vector fields $X^1, \dots, X^r$, then the mapping $T_p(M) \to \boldsymbol{R}$ defined by (5.7)

$$Y_p \to (\nabla_{Y_p}\Phi)(X_p^1, \dots, X_p^r)$$

is linear. On the other hand, it is clear that for C^∞-vector fields $X^1, \dots, X^r$; Y the function $\Psi(X^1, \dots, X^r; Y) = (\nabla_Y \Phi)(X_1, \dots, X_r)$ is C^∞. ▮

It is not difficult to give formulas in terms of the components of Φ in local coordinates for the components of Ψ. We shall give the formulas and leave their verification as an exercise. We suppose that U, φ is a local coordinate system with local coordinates $x^1, \dots, x^n$, coordinate frames $E_1, \dots, E_n$, and with $\nabla_{E_i} E_j = \sum_k \Gamma_{ij}^k E_k$. Let Φ be as above and let $\Phi_{i_1 \cdots i_r} = \Phi(E_{i_1}, \dots, E_{i_r})$ be its components.

(5.9) Corollary *With the notation above, the components*

$$\Psi_{j_1, \dots, j_{r+1}} = \Psi(E_{j_1}, \dots, E_{j_{r+1}})$$

of Ψ *on* U *are given by the formulas*

$$\Psi_{j_1, \dots, j_{r+1}} = \frac{\partial}{\partial x_{j_{r+1}}} \Phi_{j_1 \cdots j_r} - \sum_{k, i} \Gamma_{j_{r+1} j_i}^k \Phi_{j_1 \cdots k \cdots j_r}, \qquad k = 1, \dots, n; \quad i = 1, \dots, r.$$

(5.10) Definition A tensor field $\Phi \in \mathscr{T}^r(M)$ is said to be *parallel along a curve* $p(t)$ if $D\Phi/dt \equiv 0$ along the curve. It is said to be *parallel* if $D\Phi/dt = 0$ along every curve on M.

We remark that if $\nabla_{X_p}\Phi = 0$ for every $X_p \in T_p(M)$ and all $p \in M$, then it is parallel; so in fact if it is parallel along geodesics, for example, then it will be parallel. This follows from Lemma 5.5 and the fact that there is a geodesic tangent to any given vector X_p.

We also note that if $p(t)$, $a \le t \le b$, is a curve of class C^1, say, then Φ is parallel along $p(t)$ if and only if it satisfies

$$\frac{d}{dt}(\Phi(X_t^1, \dots, X_t^r)) \equiv 0$$

for every set $X_t^1, \dots, X_t^r$ of parallel vector fields along the curve $p(t)$.

(5.11) Example Let M be a Riemannian manifold of constant curvature K. Then, by definition, for any orthonormal pair of vectors X_p, Y_p the sectional curvature $R(X_p, Y_p, X_p, Y_p) = -K$. Suppose $p(t)$ is any curve through p with, say, $p(0) = p$. Let $X_{p(t)}$, $Y_{p(t)}$ be the uniquely determined parallel fields such that $X_p = X_{p(0)}$ and $Y_p = Y_{p(0)}$. Then $X_{p(t)}$, $Y_{p(t)}$ is orthonormal at each $p(t)$ and $R(X_{p(t)}, Y_{p(t)}, X_{p(t)}, Y_{p(t)}) = -K$, a constant independent of t. It follows that for any parallel vector fields along $p(t)$ say X_t^i, $i = 1, 2, 3, 4$, that

$$\frac{d}{dt} R(X_t^1, X_t^2, X_t^3, X_t^4) \equiv 0.$$

Indeed the values of all of the sectional curvatures uniquely determine the curvature; thus the curvature is parallel if it is constant on parallel sections π_t along any curve $p(t)$.

We might think that this is the only case in which the curvature tensor $R(X, Y, Z, W)$ is parallel, but in fact this is not the case as we shall now see.

(5.12) Theorem (Cartan) *If M is a Riemannian symmetric space, then the curvature tensor is parallel.*

Proof We know that any isometry of a Riemannian manifold preserves parallelism; it carries parallel vector fields, sections, and so on, along a curve to parallel vector fields, sections, and so on, along the image. Moreover isometries preserve the curvature,

$$R_p(X_p, Y_p, Z_p, W_p) = R_{F(p)}(X_{F(p)}, Y_{F(p)}, Z_{F(p)}, W_{F(p)}).$$

Finally isometries carry geodesics to geodesics. This is because each of these: parallelism, curvature, and geodesics is defined in terms of the Riemannian metric. Now to show that the curvature is parallel, it is enough to show that it is constant on parallel vector fields along geodesics. However, if $p(t)$ is a geodesic, then the vectors $X_{p(t)}$, $Y_{p(t)}$, $Z_{p(t)}$, $W_{p(t)}$ of the parallel vector field determined by $X_{p(0)}$, $Y_{p(0)}, \ldots$ are given by *isometries* τ_c of M according to Theorem VII.8.7. Therefore the curvature is constant on parallel fields along the geodesic $p(t)$, which proves the result. ∎

It is important to realize that this is more general than constant curvature. We have seen an example of a symmetric space—a compact semisimple Lie group G with bi-invariant metric—in which the curvatures on various sections π_e at the identity vary between 0 (if there is an Abelian subgroup of dimension two) and a *positive* maximum value (see Theorem 3.9). Thus G is not isotropic, hence not of constant curvature in this metric, but it does have

parallel curvature. This raises the interesting question of how those Riemannian manifolds with parallel curvature may be otherwise characterized or described. The answer to this is given by the following two theorems which will not be proved in this text but are quoted for use in the next section and for their general interest. Proofs are given by Wolf [1, pp. 30, 42].

(5.13) Theorem (Cartan) *Let M be a Riemannian manifold with parallel curvature. Then M is locally symmetric, that is, each point $p \in M$ has a neighborhood U such that there is an involutive isometry $\sigma_p: U \to U$ with p as its only fixed point.*

Of course, a manifold may be locally symmetric without being globally symmetric, that is, symmetric in the sense of our original definition of symmetric space. For example, Euclidean space or a sphere—with its usual Riemannian metric—is no longer a symmetric space if a single point is removed, since we have seen that a symmetric space is necessarily complete; but it is still locally symmetric. Even if completeness is assumed, together with parallel curvature, we still cannot be quite sure that the space is symmetric—some restrictions on the fundamental group may be involved. However, if the Riemannian manifold is complete and has parallel curvature, then we may be sure that its universal covering (with the naturally induced Riemannian metric) is a symmetric space. This is a consequence of the theorem which follows. A proof of a more general version due to Hicks is given by Wolf [1].

(5.14) Theorem (Cartan–Ambrose) *Let M and N be complete, connected Riemannian manifolds of the same dimension, each with parallel curvature, and suppose further that M is simply connected. If $p \in M$ and $q \in N$ and $\varphi: T_p(M) \to T_q(N)$ is any linear mapping which preserves the inner product and the curvature; i.e., for arbitrary X_p, Y_p, Z_p, $W_p \in T_p(M)$, we have $(\varphi(X_p), \varphi(Y_p))_q = (X_p, Y_p)_p$ and*

$$R_q(\varphi(X_p), \varphi(Y_p), \varphi(Z_p), \varphi(W_p) = R_p(X_p, Y_p, Z_p, W_p)),$$

then there is a unique C^∞ mapping $F: M \to N$ which has the properties: (i) *$F(p) = q$,* (ii) *$F_*: T_p(M) \to T_q(N)$ is the same as φ, and* (iii) *F is a Riemannian covering (that is, it is a covering such that F_* is an isometry on each tangent space—thus a local isometry).*

Exercises

1. Verify the formula of Corollary 5.9.
2. Let $\Phi(X, Y)$ be the Riemannian metric on M. Prove that $\nabla_{Z_p}\Phi = 0$ for every vector Z_p at every point $p \in M$.

3. Let σ be a field of linear operators on M, that is, an element of $\mathscr{T}^1_1(M)$, so that at each p, σ_p: $T_p(M) \to T_p(M)$ is a linear mapping. Suppose for any $X \in \mathfrak{X}(M)$ that $\sigma(X) \in \mathfrak{X}(M)$ also. Define a derivative $\nabla_{Y_p}\sigma$ for σ at a point p in the direction Y_p. Show that it will define a tensor field of type $\mathscr{T}^2_1(M)$.
4. Using Exercise 2 and the special expression in local coordinates which was given in Corollary 3.6 for R_{ijkl} on a manifold of constant curvature K, show that the curvature tensor is parallel on any such manifold.
5. Show that if an alternating covariant tensor Φ is parallel, then the exterior differential form to which it corresponds is closed. Is the converse true?
6. Let X be a C^∞-vector field on a Riemannian manifold and Φ be a tensor; for definiteness let Φ be the Riemannian metric tensor. Make a suitable definition of the Lie derivative $L_X\Phi$ and interpret $L_X\Phi = 0$.

6 Manifolds of Constant Curvature

The manifolds of constant curvature, which we have introduced briefly in previous sections, are on the one hand the simplest Riemannian manifolds and yet on the other hand are sufficiently complicated to present a fascinating object of study. Classically they are the oldest known examples in the sense that they include Euclidean space, the non-Euclidean spaces discovered by Bolyai and Lobachevskii, and the spherical and elliptic spaces, whose geometry was studied by Riemann (as examples of spaces in which no parallel geodesics existed). In fact *locally*, the geometry of any space of constant curvature is equivalent to one of these classical geometries as we shall see. What makes these spaces particularly intriguing is that questions about them can often be reduced to purely algebraic problems of an interesting nature. In the short scope of this paragraph we can only give an indication of the special flavor of this subject, but the books of Kobayashi and Nomizu [1] and Wolf [1] contain many details for the reader who wishes to go further.

We recall that a Riemannian manifold M is said to have *constant curvature* if all sectional curvatures at all points have the same constant value K. This implies that the curvature tensor is parallel and hence that the manifold is locally symmetric according to Theorem 5.13, but the converse does not hold since there are many symmetric spaces, for example any non-Abelian compact Lie group with the bi-invariant metric, which are not spaces of constant curvature. According to Corollary 3.6 and to Theorem 4.9, or at least to its proof, it should be possible to give a characterization of Riemannian manifolds of constant curvature in terms of differential forms. To this end we suppose M to be a Riemannian manifold and let ω^i, $1 \le i \le n$, denote the field of coframes dual to an orthonormal frame field $E_1, \ldots, E_n$ on

an open set $U \subset M$, with ω_i^j, $1 \leq i, j \leq n$, denoting the corresponding connection forms. We then state the following lemma, whose proof is contained in Theorem 4.9 and Exercise 1.

(6.1) Lemma *If M has constant curvature K, then the curvature forms $\Omega_i^j = d\omega_i^j - \sum_k \omega_i^k \wedge \omega_k^j$ are given by*

$$\Omega_i^j = -K\omega^i \wedge \omega^j.$$

Conversely, if on a neighborhood U of each point of M there is an orthonormal frame field $E_1, \ldots, E_n$ for which the uniquely determined ω^i, ω_i^j satisfy this equation, then M has constant curvature K.

We shall use this presently to give an example of a manifold of constant negative curvature. Before doing so we recall that Euclidean space with its standard Riemannian metric is a space of zero curvature and that the n-sphere of radius a in $\boldsymbol{R}^{n+1}$ with the induced Riemannian metric has constant curvature $K = 1/a^2$. Thus for every nonnegative real number K, we have already found an example of Riemannian manifold of arbitrary dimension n with constant curvature K. We shall next give an example of an n-dimensional Riemannian manifold of constant curvature $K = -1$. A slight variation (Exercise 2) will produce an example for any $K < 0$.

(6.2) Example (*Hyperbolic space*) Let M be the open upper half-space of $\boldsymbol{R}^n$ defined by $M = \{x \in \boldsymbol{R}^n \mid x^n > 0\}$ with the Riemannian metric given by the line element (see Section V.3)

$$ds^2 = \frac{(dx^1)^2 + \cdots + (dx^n)^2}{(x^n)^2}.$$

More precisely, we note that, as a manifold, M is covered by a single coordinate system with local coordinates $x^1, \ldots, x^n$ and coordinate frames $\partial/\partial x^1, \ldots, \partial/\partial x^n$. This is because, as a manifold, M corresponds to an open subset of $\boldsymbol{R}^n$. In these local coordinates, the components of the Riemannian metric Φ are given by

$$g_{ij}(x) = \Phi\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{\delta_{ij}}{(x^n)^2}.$$

We use Lemma 6.1 to see that this manifold has constant curvature $K = -1$ as claimed. (When $n = 2$, this is the Riemannian manifold of Example VII.9.5.)

Let $E_i = x^n(\partial/\partial x^i)$, $i = 1, \ldots, n$; these define an orthonormal frame field on all of M. We denote by $\omega^1, \ldots, \omega^n$ the dual coframes which are given by

$\omega^i = (1/x^n)\, dx^i$, $i = 1, \ldots, n$. It is easy to verify that the forms $\omega_i^j = \delta_{nj}\omega^i - \delta_{ni}\omega^j$ satisfy the equations

$$d\omega^i = \sum_{j=1}^{n} \omega^j \wedge \omega_j^i \qquad \text{and} \qquad \omega_i^j + \omega_j^i = 0;$$

hence they must be the connection forms since these are uniquely determined by these conditions. Finally, taking the exterior derivative of ω_i^j we obtain

$$\Omega_i^j = d\omega_i^j - \sum \omega_i^k \wedge \omega_k^j = +\omega^i \wedge \omega^j.$$

Then from Lemma 6.1 it follows that M has constant curvature $K = -1$. We call this *hyperbolic space* and denote it by H^n (for its underlying space, the "half-plane").

Thus we have examples of spaces of positive, zero, and negative constant curvature. Note that all three examples are simply connected: when $K > 0$, our example was the compact manifold S^n and, when $K = 0$ or $K = -1$, the corresponding manifolds $\boldsymbol{E}^n$ and H^n are diffeomorphic to $\boldsymbol{R}^n$. For convenience, in what follows we suppose that S^n has radius $+1$ and hence $K = +1$. Since S^n is compact, it is complete; we also know $\boldsymbol{E}^n$ to be a complete Riemannian manifold; and we shall prove later that H^n is complete. The importance of these facts stem from the following theorem (which goes back to much earlier work of Killing and Hopf).

(6.3) Theorem *Every complete, simply connected Riemannian manifold M of constant curvature $K = +1, 0,$ or -1 is isometric to one of the three examples above, in fact to S^n if $K = +1$, to $\boldsymbol{E}^n$ if $K = 0$, and to H^n if $K = -1$. More precisely, given $p \in M$, and q in either S^n, $\boldsymbol{E}^n$, or H^n according to whether $K = +1$, 0, or -1, and given a prescribed linear map of $T_p(M)$ onto the tangent space at q which preserves the inner product, then there is exactly one isometry F of M to the corresponding space of constant curvature taking p to q and such that F_* corresponds to the given linear mapping on $T_p(M)$.*

This is an immediate consequence of Theorem 5.14 once we know that H^n is complete—which is proved later. We remark that using Exercise 2 and spheres S^n of arbitrary radius, we may extend this theorem without difficulty to spaces of any constant curvature. Although Theorem 5.14 is not proved in this text, this consequence of it will be used. We have the following obvious corollary of Theorem 6.3.

(6.4) Corollary *Let M be S^n, $\boldsymbol{E}^n$, or H^n and let $E_{1p}, \ldots, E_{np}$, $E_{1q}, \ldots, E_{nq}$ be orthonormal frames at two arbitrary points p, q of M. Then there is a unique isometry of M taking p to q and E_{ip} to E_{iq}, $i = 1, \ldots, n$.*

This shows that the group of isometries is transitive on M and makes it plausible that in each of these cases it is a Lie group. We already know this however for S^n, whose group of isometries is $O(n + 1)$ and for E^n, whose group of isometries consists of rotations and translations and their products (Example III.7.6). The group of all isometries of H^n will be studied only in the special case $n = 2$, although some indication will be given of the general situation.

In order to carry our study of these spaces somewhat further we make some comments concerning covering spaces. If M is a Riemannian manifold and $\tilde{M}$ a covering manifold with covering mapping $F: \tilde{M} \to M$, then there is a unique Riemannian metric on $\tilde{M}$ such that F is a local isometry. When $\tilde{M}$ has this metric, the covering will be called a *Riemannian covering*. The following facts are quite easily verified from the definitions: (i) F carries geodesics to geodesics and each geodesic on M is covered by a unique geodesic on $\tilde{M}$; (ii) If M is complete, then $\tilde{M}$ is also complete (convergence of Cauchy sequences is a local phenomenon); and (iii) the covering transformations are isometries of $\tilde{M}$. With the aid of these facts one may reduce the determination of manifolds of constant curvature to a group theoretic problem—or at least make the first step in that direction.

(6.5) Theorem *Let M be a complete manifold of constant curvature $K = +1, 0,$ or -1. Then the universal covering manifold $\tilde{M}$ is isometric to S^n, E^n, or H^n, respectively, and M is the orbit space of a subgroup Γ of the group of isometries of $\tilde{M}$ which acts freely and properly discontinuously on $\tilde{M}$.*

The theorem follows from the fact that $\tilde{M}$ is complete, simply connected, and (since the covering mapping is a local isometry) has the same constant curvature as M. We know from the theory of covering spaces that $M = \tilde{M}/\Gamma$ and that the covering transformations Γ act freely and properly discontinuously (as a group of isometries). We give some indication of how this may be used by considering some examples.

Spaces of Positive Curvature

In order to find Riemannian manifolds of constant positive curvature $K = +1$, it is necessary to find subgroups Γ of the group of isometries of S^n, the unit sphere, which act freely and properly discontinuously on S^n. The isometries of S^n are contained in $O(n + 1)$ which acts in the usual way on the unit sphere in $\boldsymbol{R}^{n+1}$, hence $\Gamma \subset O(n + 1)$. The assumption that Γ acts freely means that no element of Γ, except the identity, leaves a point of S^n fixed. Thus if $A \in \Gamma$ and $A \neq I$, A cannot have $+1$ as a characteristic value. Moreover, Γ must be a group of finite order, since otherwise there must be an $x \in S^n$ such that $\Gamma x = \{Ax \mid A \in \Gamma\}$ has a limit point, which would contradict the proper discontinuity. Thus we must find finite subgroups of

$O(n+1)$ no element of which (except the identity) leaves a vector x fixed. This is clearly a necessary condition for Γ, it is also sufficient (Exercise 3).

The simplest example of a subgroup Γ of $O(n+1)$ of the type described is the group consisting of two elements, $\Gamma = \{\pm I\}$. The orbit space S^n/Γ is the collection of all antipodal pairs of points on S^n and is, as we have mentioned earlier, just the projective space $P^n(\boldsymbol{R})$ (Example III.2.5 and Exercise III.2.3). Thus for every n we have at least two inequivalent spaces of constant curvature—real projective space and its universal (Riemannian) covering space S^n. When n is even, we have the following fact.

(6.6) *If n is even, then S^n and $P^n(\boldsymbol{R})$ are the only complete manifolds of constant curvature $K = +1$.*

Proof This is seen as follows. Let Γ be a properly discontinuous group of isometries acting freely on S^n. Then $\Gamma \subset O(n+1)$ and each $A \in \Gamma$ is an $(n+1) \times (n+1)$ orthogonal matrix. Therefore A must have a real characteristic value since the degree of its characteristic polynomial is an odd number $n+1$. Since the characteristic values of an orthogonal matrix are of absolute value one, A has ± 1 as a characteristic value. We have seen that only the identity on Γ can have $+1$ as a characteristic value, hence -1 is a characteristic value of A. This implies that A^2 has $+1$ as characteristic value, so $A^2 = I$. This means that each of the characteristic values of A is either $+1$ or -1, and hence, either all are $+1$ and $A = I$, or all are -1 and $A = -I$. This completes the proof when combined with the example mentioned above. ∎

(6.7) Example When n is odd, other possibilities can occur. As an indication we will show that in the case of S^3 there exist many examples of finite subgroups $\Gamma \subset O(4)$ which act freely on S^3 and thus give manifolds S^3/Γ of constant positive curvature. The examples are based on the algebra $\boldsymbol{K}$ of quaternions, that is, on the real linear combinations

$$\mathbf{q} = x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}$$

of the four symbols 1, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ with the usual rules of multiplication and with componentwise addition (see Chevalley [1]). We denote by $\bar{\mathbf{q}}$, the conjugate of $\mathbf{q}$,

$$\bar{\mathbf{q}} = x - y\mathbf{i} - z\mathbf{j} - w\mathbf{k}$$

and by $\|\mathbf{q}\|$ the usual norm $\|\mathbf{q}\| = (\mathbf{q}\bar{\mathbf{q}})^{1/2}$. Then $\boldsymbol{K}$ is in obvious one-to-one linear correspondence with $\boldsymbol{R}^4$ and this corresponds to the standard norm in $\boldsymbol{R}^4$. Hence $\boldsymbol{K}_1 = \{\mathbf{q} \mid \|\mathbf{q}\| = 1\}$, the quaternions of norm one, correspond to $S^3 \subset \boldsymbol{R}^4$. As usual we identify $\boldsymbol{K}$ and $\boldsymbol{R}^4$ as vector spaces and as manifolds and we identify $\boldsymbol{K}_1$ and S^3 as manifolds. The important thing for us is that $\boldsymbol{K}_1$ is a group with respect to quaternion multiplication, since $\|\boldsymbol{q}_1\boldsymbol{q}_2\| = \|\boldsymbol{q}_1\|\|\boldsymbol{q}_2\|$. Thus, if $\boldsymbol{q} \in \boldsymbol{K}_1$, then left translation $L_{\mathbf{q}}\colon \boldsymbol{K} \to \boldsymbol{K}$ defined by

$L_{\mathbf{q}}(\mathbf{x}) = \mathbf{q}\mathbf{x}$ is an $\boldsymbol{R}$-linear mapping of $\boldsymbol{K}$ onto $\boldsymbol{K}$ and preserves the norm of $\mathbf{x}$, that is, $\|L_{\mathbf{q}}(\mathbf{x})\| = \|\mathbf{x}\|$. This means that as a linear transformation of $\boldsymbol{K} = \boldsymbol{R}^4$, $L_{\mathbf{q}}$ is an orthogonal transformation. In brief, $S^3 = \boldsymbol{K}_1$ is a group space and left translations are orthogonal transformations, in fact isometries, of S^3 with its usual Riemannian structure. Since no left translation except the identity can have a fixed point, we need only find examples of finite subgroups Γ of $\boldsymbol{K}_1$—each such example determines a three-dimensional manifold of constant positive curvature and they are all determined this way.

To find finite subgroups of $\boldsymbol{K}_1$ one uses the following fact (Exercise 4). There is a natural homorphism $\pi\colon \boldsymbol{K}_1 \to SO(3)$ which is onto and has kernel ± 1 ($+1$ is the unit quaternion). This homomorphism is given as follows: Let $\boldsymbol{R}^3$ be identified with the subspace of $\boldsymbol{K}$ of all quaternions of the form $\mathbf{q} = y\mathbf{i} + z\mathbf{j} + w\mathbf{k}$, with real part $x = 0$. Then to each $\mathbf{q}' \in \boldsymbol{K}_1$ we let correspond the rotation $\pi(\mathbf{q}')$ of $\boldsymbol{R}^3$ given by $\mathbf{q} \to \mathbf{q}'\mathbf{q}(\mathbf{q}')^{-1}$. Now given any finite subgroup $\Gamma_1 \subset \mathrm{SO}(3)$, then $\Gamma = \pi^{-1}(\Gamma_1)$ is a finite subgroup of $\boldsymbol{K}_1$. Such subgroups of $SO(3)$ are easy to find—the group of symmetries of any regular solid (omitting those of determinant -1) give examples. The problem of classifying *all* complete spaces of constant positive curvature has recently been completed by Wolf [1]; classification of the finite subgroups of $\boldsymbol{K}_1$ and of $SO(3)$ is carried out as an example on p. 83ff of his book.

Spaces of Zero Curvature

Now consider the Riemannian manifolds which have Euclidean space of the same dimension as their universal Riemannian covering space; they are the (complete) spaces of zero curvature. Thus they are of the form $M = \boldsymbol{E}^n/\Gamma$, the orbit space of a subgroup Γ of the group of isometries (rigid motions) of $\boldsymbol{E}^n$. If we identify $\boldsymbol{E}^n$ with $\boldsymbol{R}^n$ and use vector space notation, then each isometry is of the form $x \to Ax + b$, where $A \in O(n)$ and $b = (b^1, \ldots, b^n)$, and determine, respectively, a rotation and translation of the space (Examples III.7.6 and IV.9.4). Since, locally at least, the geometry of any such M is just that of Euclidean space, these spaces might seem to lack interest. This is not the case however; in particular, the global behavior of geodesics is very different from that of geodesics in $\boldsymbol{E}^n$. We have already noted this in the case of two examples: the cylinder, which is just $\boldsymbol{E}^2/\Gamma$ with $\Gamma = \{x \to x + ne_1 \mid e_1 = (1, 0),\ n \in \boldsymbol{Z}\}$, and the torus T^2 obtained as the orbit space of the group of translations $\{x \to x + ne_1 + me_2 \mid n, m \in \boldsymbol{Z},\ e_1 = (1, 0),\ e_2 = (0, 1)\}$.

Historically the study of these spaces is closely linked to that of the study of crystal structures on the plane $\boldsymbol{E}^2$ and in Euclidean space $\boldsymbol{E}^3$, that is, to uniform coverings of the plane by congruent polygons and of $\boldsymbol{E}^3$ by congruent polyhedra. It is fairly easy to convince ourselves that the symmetries of

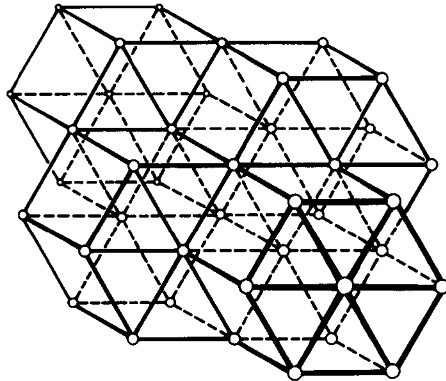


Figure VIII.9

such crystalline structures—rigid motions carrying the structures onto themselves—form a subgroup Γ of the group of rigid motions which acts properly discontinuously (Fig. VIII.9). Elements of such groups may well have fixed points however, so these groups are somewhat more general than those which generate examples of manifolds of zero curvature. It was proved in the 19th century by several mathematicians independently (by classification of all possibilities) that there were only a finite number of crystal structures on E^3. This gave rise to the question posed by Hilbert [2] in his famous address of 1900 as to whether the number of possible isomorphism classes of properly discontinuous groups of motions Γ of E^n for which the orbit space E^n/Γ is compact is finite for every n. These are called crystallographic groups, and Hilbert's question was answered affirmatively by Bieberbach [1] in 1911. This implies, in particular, that for every dimension n there exist at most a finite number of compact Riemannian manifolds of curvature zero. Among these, of course, is the torus T^n, and it is a consequence of Bieberbach's work that every such manifold has the torus as covering space. The proof of these theorems involves very interesting group theoretic arguments; it may be found in the books of Kobayashi and Nomizu [1] and Wolf [1]. In the latter book a complete classification of the manifolds of zero curvature in dimensions 2 and 3 is given; no general classification for all n is known. (See also Grove and Benson [1].)

Spaces of Constant Negative Curvature

First we consider H^2 as given in Example 6.2, except that we shall write (x, y) for (x^1, x^2) and identify H^2 with the upper half-plane of the complex numbers C by the correspondence $(x, y) \leftrightarrow z = x + iy$. Then H^2 is the open

subset of C, consisting of all complex numbers z with positive imaginary part $\operatorname{Im} z > 0$. We may then write the Riemannian metric, or line element $ds^2 = \sum_{i,j=1}^{2} g_{ij}\, dx^i\, dx^j$, in the complex or real form

$$ds^2 = \frac{dz\, d\bar{z}}{(\operatorname{Im} z)^2} = \frac{dx^2 + dy^2}{y^2}.$$

We have already considered this Riemannian manifold and its isometries (Example VII.9.5). The reason for passing to complex coordinates is that it makes it much simpler to define and work with the group of isometries. Of course, other representations of H^2 and its group of isometries are often used—some of which extend to H^n for all n, but the technical difficulties would be greater for us at this stage (see Wolf [1, Section 2.4]). Recall that mappings on C of the form $z \mapsto w = (az + b)/(cz + d)$, $a, b, c, d \in C$ such that $ad - bc \neq 0$, are isometries of H^2; in analytic function theory they are called *linear fractional transformations* (see Ahlfors [1] for example). The following theorem restates Example VII.9.5 and adds a little to it.

(6.8) Theorem *The group G of linear fractional transformations such that a, b, c, d are real numbers and $ad - bc = +1$ is exactly the group of isometries of H^2 identified with the upper half-plane of C. The mapping $F: Sl(2, \mathbf{R}) \to G$ defined by letting the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ correspond to the linear fractional transformation $z \mapsto w = (az + b)/(cz + d)$ is a homomorphism of $Sl(2, \mathbf{R})$ onto G with kernel $\pm I$.*

Proof Except for the assertion that this group contains *all* of the isometries, these statements were all proved, or given as problems, in Example VII.9.5. To review briefly the arguments, which the reader should check in detail, we note that the last statement is verified by a straightforward computation. Whereas to see that the first statement is correct, we note that if w is the image of $z \in H^2$ by a transformation of G, then

$$\operatorname{Im} w = \frac{\operatorname{Im} z}{|cz + d|^2} > 0$$

so that the upper half-plane maps onto itself. If we compute dw, we find that

$$dw = \frac{dz}{(cz + d)^2}$$

from which it follows that

$$\frac{dw\, d\bar{w}}{(\operatorname{Im} w)^2} = \frac{dz\, d\bar{z}}{(\operatorname{Im} z)^2}$$

so that ds^2 is preserved—a shorthand way of seeing that the components of g_{ij} transform as they should for an isometry. [Of course, this mapping could

be given in terms of real and imaginary parts, that is, the functions, $u(x, y)$ and $v(x, y)$, such that $w = u(x, y) + iv(x, y)$ could be computed and the mapping written without use of complex variables; but the computations become much more difficult.] In order to see that this group G contains *all* isometries, we recall first that it acts transitively on the upper half-plane and second that it is transitive on directions. Indeed, in the example cited it was shown that the orbit of $i = \sqrt{-1}$ is all of H^2, which implies transitivity and that the isotropy subgroup of i consists of elements of G corresponding to matrices in $Sl(2, \boldsymbol{R})$ of the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

This subgroup of G is transitive on directions at i; in fact it acts as $SO(2)$ on the tangent space to H^2 at i. These facts together with Corollary 6.4 prove the assertion. ∎

We note that angles on H^2 in terms of the given Riemannian metric are the same as angles on $\boldsymbol{R}^2$, moreover—as is well known—linear fractional transformations are analytic mappings on the complex plane and as such are conformal, that is, they preserve angles between curves. We will also use from complex function theory the fact that linear fractional transformations carry circles and straight lines of $\boldsymbol{C}$ into circles and straight lines (see Ahlfors [1]). Thus any circle which is orthogonal to the real axis will be carried by any element of G into a circle orthogonal to the real axis or a vertical straight line. We have left it as an exercise to prove that vertical straight lines are geodesics of H^2. Then it follows rather easily that any circle orthogonal to the real axis is also a geodesic. In fact a little simple Euclidean geometry (Fig. VIII.10) shows that through a given $z_0 \in H^2$ there is exactly one circle (or vertical line) tangent to each direction at z_0 and orthogonal to

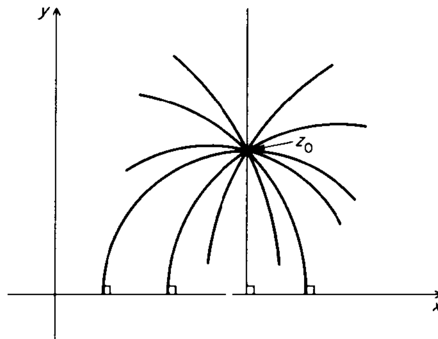


Figure VIII.10

the real axis. Since isometries take geodesics to geodesics, this gives every geodesic through z_0. One important consequence is that every geodesic can be extended to infinite length so that H^2 is seen to be a complete metric space. It is sufficient to check this for just one geodesic, namely, $x = 0$, $y = t$, $0 < t < \infty$. The length of this geodesic from $t = a$ to $t = b$ is $\int_a^b dt/t$ so it is unbounded in both directions, that is as $a \to 0$ or $b \to \infty$, which shows it is indefinitely extendable. We also saw in Section VII.9 that H^2 is an example of a symmetric space, which means that it must be complete (Theorem VII.8.4). We have previously noted that H^2 is the space of non-Euclidean geometry and that it is easy to see from this description of geodesics that Euclid's postulate of parallels is not satisfied (although all the other postulates of Euclid are!). This is illustrated in Fig. VII.15. This behavior of geodesics should be contrasted with that on S^2 and $P^2(\boldsymbol{R})$, spaces of constant *positive* curvature on which every pair of geodesics intersect—twice on S^2 and once $P^2(\boldsymbol{R})$.

We turn now to consideration of H^n, $n > 2$ as described in Example 6.2 and use this information to verify that H^n is complete. First we note that any translation of H^n in a direction parallel to the plane $x^n = 0$ is an isometry. The same holds for a rotation of the underlying $\boldsymbol{R}^n$ which leaves x^n fixed, in other words, a linear transformation of the variables $x^1, \ldots, x^{n-1}$ with orthogonal matrix is an isometry. Thus any 2-plane determined by a point $x \in H^n$ and unit vector X_x at x can be carried to the submanifold $H^2 = \{x \in H^n \mid x^1 = \cdots x^{n-1} = 0\}$ by an isometry of H^n. If we then verify that geodesics of H^2 are geodesics of H^n (Exercise 10), it will follow from the above facts concerning H^2 and known properties of geodesics that every geodesic of H^n can be extended to infinite length. This means that H^n is complete; it also means that the geodesics of H^n are exactly the semicircles whose center lies on the $(n-1)$-plane $x^n = 0$ and whose plane is perpendicular to it.

The geometry of H^2 is extremely useful in analytic function theory and the subgroups Γ of G which operate properly discontinuously on H^2 are extensively studied in automorphic function theory (see Lehner [1] and Siegel [1]). In fact automorphic functions are precisely those complex analytic functions on H^2 whose value is the same at each point of the orbit of some such Γ. Thus they define functions on H^2/Γ, the space of orbits. This is analogous to doubly periodic functions on $\boldsymbol{C}$ which take the same value at each point of the orbit of a group Γ of the form $\Gamma = \{z \to z + mw_1 + nw_2\}$ for (independent) $w_1, w_2 \in \boldsymbol{C}$ and thus define functions on $\boldsymbol{C}/\Gamma = T^2$. The best known automorphic function is the one associated with the subgroup of G for which a, b, c, d are integers in each linear fractional transformation, that is, the image of $Sl(2, \boldsymbol{Z}) \subset Sl(2, \boldsymbol{R})$ under the homomorphism of Theorem 6.8. This group is known as the *elliptic modular group*. It acts discontinuously on H^2 but some elements have fixed points; however, it

contains subgroups which act freely and thus determine a manifold H^2/Γ of curvature $K = -1$. Using analytic function theory or the geometry of H^2, it is possible to show that there exist subgroups Γ of G acting freely and properly discontinuously so that H^2/Γ is a compact manifold; in fact every surface of genus $g > 1$ can be obtained in this manner and hence every such surface has a Riemannian metric for which the Gaussian curvature is constant and equal to -1. For $n > 2$, it is much more difficult to find subgroups Γ of the group of isometries of H^n such that H^n/Γ is a compact manifold; in fact this is an area of active research at present and has many interesting unsolved problems.

Exercises

1. Prove Lemma 6.1 in detail, using the known facts about curvature forms and equations of structure.
2. Let a be any positive real number and let M be the subspace of $\boldsymbol{R}^n$ such that $x^n > 0$. Then the Riemannian metric on M given by $g_{ij}(x) = (a^2/(x^n)^2)\,\delta_{ij}$ has constant curvature $K = -1/a^2$.
3. Show that if Γ is a finite subgroup of $O(n+1)$, which has the property that no $A \in \Gamma$ except I, the identity, has $+1$ as a characteristic value, then Γ acts freely and properly discontinuously on S^n.
4. Show that a discontinuous group of isometries of a Riemannian manifold is necessarily properly discontinuous.
5. Show that the homomorphism $\pi\colon \boldsymbol{K}_1 \to SO(3)$ of Example 6.7 is indeed a homomorphism onto $SO(3)$ with kernel ± 1 as claimed.
6. Let G be a connected Lie group and K a compact subgroup, and suppose that G acts on G/K by left translation (in the usual way). Show that a subgroup Γ of G acts properly discontinuously on G/K if and only if it is discrete. Show that if Γ has no elements of finite order, then it acts freely. Apply this to G, the group of motions of $\boldsymbol{E}^n$ [with $K = 0(n)$].
7. Does a rigid motion $x \to Ax + b$, $A \in O(3)$ and $b = (b^1, b^2, b^3)$, of the space $\boldsymbol{E}^3$ identified with $\boldsymbol{R}^3$ have a fixed point?
8. Show that the subgroup Γ of rigid motions of $\boldsymbol{R}^3$ generated by translations $x \to m\mathbf{e}_1 + n\mathbf{e}_2 + p\mathbf{e}_3$, $m, n, p \in \boldsymbol{Z}$ and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, the standard basis, together with the motion $x \to Ax + \frac{1}{2}\mathbf{e}_1$, where $A(\mathbf{e}_1) = \mathbf{e}_1$, $A(\mathbf{e}_2) = -\mathbf{e}_2$, and $A(\mathbf{e}_3) = -\mathbf{e}_3$, acts freely and properly discontinuously on $\boldsymbol{R}^3$ and that $\boldsymbol{R}^3/\Gamma$ is compact.
9. Show that vertical lines $x =$ constant on the space H^2 are geodesics.
10. Compute Γ^k_{ij} for H^n [using the natural coordinates $(x^1, \ldots, x^n)$] and show that a geodesic of the submanifold $H^2 = \{x \in H^n \mid x^1 = \cdots = x^{n-2} = 0\}$ is a geodesic of H^n.

11. Show that the group Γ of linear fractional transformations of the form $w = (az + b)/(cz + d)$, a, b, c, d, integers such that $ad - bc = 1$, acts properly discontinuously on H^2 but that it does not act freely.
12. Let K be a real number and let $\rho = 1 + (K/4) \sum_{i=1}^{n} (x^i)^2$. Prove that if a Riemannian metric is given on a coordinate neighborhood U, φ of an n-dimensional manifold M, $\varphi(U) = B_\varepsilon^n(0)$, (for some $\varepsilon > 0$) by

$$g_{ij}(x) = \frac{1}{\rho^2} \delta_{ij},$$

then on U this metric has constant sectional curvature K.

Notes

The ideas touched upon in this chapter and the previous one span the entire history of differential geometry, from the early work of Gauss [1] and Riemann [1] through that of Cartan [1] right down to the present. Given the scope of the subject, its treatment here was necessarily both selective and brief. However, for readers who wish to go further into some of the topics we have touched upon in Chapters VII and VIII, there are many excellent books available, of which we shall mention several. For surface theory Stoker [1], O'Neill [1], do Carmo [1], and Millman and Parker [1] are very helpful. These books are geometric and intuitive in approach yet lead directly toward the current work in manifolds of arbitrary dimension, whereas many other books on "classical" differential geometry do not. For the reader who wishes to delve more deeply into the subject of spaces of constant curvature, the book by Wolf [1] is an excellent source, especially for the zero and positive curvature cases. It also contains a very complete bibliography. A good introduction to the sort of problems one will encounter in spaces of negative curvature may be found in such books as those by Lehner [1] and Siegel [1], which deal exhaustively with the two dimensional case and its relations to Riemann surfaces and automorphic function theory. For Riemannian Geometry in general, the encyclopedic two volume work of Kobayashi and Nomizu [1] contains a wealth of information and a very complete bibliography. For questions concerning symmetric spaces, the reader is referred to Helgason [1], which also has an extensive bibliography. These, together with Milnor [1], Klingenberg [1], and Cheeger and Ebin [1] will give some idea of the current thrust of the theory and of its richness and diversity.

Most of the current interest in Riemannian geometry is in what are known as *global* problems, which in very many cases are concerned with the relation of (often purely local) properties of the curvature of the Riemannian metric of the manifold to its global geometric structure, for example, to its topology, Euler characteristic, and so on. As an epilogue to this chapter we shall mention several famous results along these lines which we did not have the time or space to take up although they are easily accessible to the reader at this point. There are a number of results which draw conclusions about the manifold from the assumption that it has a Riemannian metric whose sectional curvatures are all of the same sign—but not necessarily constant. For example, if they are all greater than a positive constant ε, then the manifold is compact. Since its universal Riemannian covering manifold necessarily has the same property, it must also be compact, from which we can conclude that the fundamental group of the original manifold is finite. On the other hand, it has been shown that if the sectional curvatures of a Riemannian manifold are all negative, then the universal covering manifold must be diffeomorphic to $\boldsymbol{R}^n$, a fact which has strong implications for the deRham groups of the manifold. Many beautiful results of this type may be found in Milnor's book [1], which is highly recommended

for further reading. It is not evident from the two examples cited, but it is a fact that the influence of the curvature on the structure of the geodesics is crucial to many such results. In addition to Milnor's book, the reader will find many interesting results of this sort—and with less emphasis on topology—given by Bishop and Crittenden [1].

As a final example we mention the famous classical theorem of Gauss and Bonnet which gives the following relation between the Gaussian curvature K and Euler characteristic χ of a compact orientable surface M:

$$2\pi\chi = \int_M K\,dA \qquad (dA = \text{area element on } M).$$

This has many interesting consequences. For example, if M has a Riemannian metric such that $K > 0$ everywhere, then it is homeomorphic to S^2, and if the metric is such that $K < 0$ everywhere, then it must have genus $g > 1$, that is, it must be homeomorphic to a sphere with two or more handles attached. A proof of this theorem is given in the books on surface theory referred to above. The generalization of the theorem to higher dimensions is not easy and requires some use of algebraic topology. In fact, this theorem resisted generalization for many decades and its extension to higher dimensional Riemannian manifolds by Allendoerfer–Weil and by Chern [3, 4]—especially Chern's method of proof—led to many new problems in differential geometry and to the discovery of further important relations between the Riemannian geometry and the topology of manifolds.

REFERENCES

ABRAHAM, R., and MARSDEN, J. E.
[1] "Foundations of Mechanics," 2nd Ed. Benjamin/Cummings, Reading, Massachusetts, 1978.

ABRAHAM, R., MARSDEN, J. E., and RATIU, T.
[1] "Manifolds, Tensor Analysis, and Applications." Addison Wesley, 1983.

ADAMS, J. F.
[1] "Lectures on Lie Groups." Benjamin, New York, 1969.

AHLFORS, L. V.
[1] "Complex Analysis." McGraw-Hill, New York, 1953.

APOSTOL, T. M.
[1] "Mathematical Analysis." Addison-Wesley, Reading, Massachusetts, 1957.

ARNOLD, V. I.
[1] "Ordinary Differential Equations." MIT Press, Cambridge, Massachusetts, 1978.
[2] "Mathematical Methods of Classical Mechanics." Springer-Verlag, New York, 1978.

AUSLANDER, L., AND MACKENZIE, R. E.
[1] "Introduction to Differentiable Manifolds." McGraw-Hill, New York, 1963.

BIEBERBACH, L.
[1] Über die Bewegungsgruppen der Euklidischen Räume I, Math. Ann. **70** (1911), 271–336; II, **72** (1912), 400–412.

BISHOP, R. L., AND CRITTENDEN, R. J.
[1] "Geometry of Manifolds." Academic Press, New York, 1964.

BOHR, H.
[1] "Collected Mathematical Works." Dansk Matematisk Forening, Copenhagen, 1952.

BOOTHBY, W. M.
[1] On two classical theorems of algebraic topology, *Amer. Math. Month.* **78** (1971), 237–249.

BOTT, R., and TU, L. W.
[1] "Differential Forms in Algebraic Topology." Springer-Verlag, New York, 1982.

BROCKER, T., and JANICH, K.
[1] "Introduction to Differential Topology." Cambridge Univ. Press, Cambridge, England, 1982.

CARTAN, É.
[1] "Oeuvres Complètes." Gauthier-Villars, Paris, 1952.

CHEEGER, J., and EBIN, D.
[1] "Comparison Theorems in Riemannian Geometry," North-Holland Publ., Amsterdam, 1975.

CHERN, S. S.
[1] Differentiable Manifolds, Univ. of Chicago lecture notes, 1952.
[2] The geometry of G-structures, *Bull. Amer. Math. Soc.* **72** (1966), 167–219.
[3] A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds, *Ann. of Math.* **45** (1944), 747–752.
[4] Some new viewpoints in differential geometry in the large, *Bull. Amer. Math. Soc.* **52** (1946), 1–30.
[5] "Curves and Surfaces in Euclidean Space," Studies in Global Geometry and Analysis, Prentice Hall, Englewood Cliffs, New Jersey, 1967.

CHERN, S. S., AND CHEVALLEY, C.
[1] Elie Cartan and his mathematical work, *Bull. Amer. Math. Soc.* **58** (1952), 217–250.

CHEVALLEY, C.
[1] "Theory of Lie Groups." Princeton Univ. Press, Princeton, New Jersey, 1946.

CHEVALLEY, C., AND EILENBERG, S.
[1] Cohomology theory of Lie groups and Lie algebras, *Trans. Amer. Math. Soc.* **63**, (1948), 85–124.

CURTIS, M. L.
[1] "Matrix Groups," 2nd Ed., Springer-Verlag, New York, 1984.

DIEUDONNÉ, J.
[1] "Foundations of Modern Analysis." Academic Press, New York, 1960.

DO CARMO, M. P.
[1] "Differential Geometry of Curves and Surfaces." Prentice Hall, Englewood Cliffs, New Jersey, 1976.

DUGUNDJI, J.
[1] "Topology," Allyn & Bacon, Boston, Massachusetts, 1966.

FLANDERS, H.
[1] "Differential Forms." Academic Press, New York, 1963.

FLEMING, W. H.
[1] "Functions of Several Variables." Addison-Wesley, Reading, Massachusetts, 1965.

FREEDMAN, M. H.
[1] The topology of four-dimensional manifolds, *J. Diff. Geom.* **17** (1982), 357–343.

GALOIS, E.
[1] "Écrits et Mémoires Mathématiques." Gauthier-Villars, Paris, 1962.

GAUSS, C. F.
[1] "General Investigations of Curved Surfaces." Raven Press, Hewlett, New York, 1965.

GODBILLON, C.
[1] "Géométrie Différentielle et Mécanique Analytique." Hermann, Paris, 1969.

GOMPF, R. E.
[1] Three exotic R's and other anomalies, *J. Diff. Geom.* **18** (1983), 317-328.

GREENBERG, M. J., and HARPER, J. R.
[1] "Algebraic Topology: A First Course," Benjamin, New York, 1981.

GREUB, W., HALPERIN, S., AND VANSTONE, R.
[1] "Connections, Curvature and Cohomology," Vol. 1. Academic Press, New York, 1972.

GROVE, L. C., and BENSON, C. T.
[1] "Finite Reflection Groups," 2nd Ed., Springer-Verlag, New York, 1985.

GUILLEMIN, V., and POLLACK, A.
[1] "Differential Topology." Prentice Hall, Englewood Cliffs, New Jersey, 1974.

HAEFLIGER, A.
[1] Varietes feuilletees, *Ann. Scuola Norm. Pisa* **3** (1962), 367-397.

HELGASON, S.
[1] "Differential Geometry Lie Groups, and Symmetric Spaces," Academic Press, New York, 1978.

HILBERT, D.
[1] "The Foundations of Geometry." (Translated by E. J. Townsend), Open Court Publ. Co., LaSalle, Illinois, 1965.
[2] Mathematical Problems, *Bull. Amer. Math. Soc.* **8** (1902), 437-479.

HILBERT, D., AND COHN-VOSSEN, S.
[1] "Geometry and the Imagination." Chelsea, New York, 1952.

HIRSCH, M. W.
[1] "Differential Topology," Springer-Verlag, New York, 1976.

HIRSCH, M. W., AND SMALE, S.
[1] "Differential Equations, Linear Algebra, and Dynamical Systems." Academic Press, New York. 1974.

HOCHSCHILD, G.
[1] "The Structure of Lie Groups." Holden-Day, San Francisco, 1965.

HOFFMAN, K., AND KUNZE, R.
[1] "Linear Algebra." Prentice-Hall, Englewood Cliffs, New Jersey. 1961.

HOPF, H.
[1] Zum Clifford-Kleinschen Raum Problem, *Math. Ann.* **95** (1925), 313-339.

HOPF, H., AND RINOW, W.
[1] Über den Begriff der vollständigen differentialgeometrischen Flächen, *Comment. Math. Helv.* **3** (1931), 209-225.

HUREWICZ, W.
[1] "Lectures in Ordinary Differential Equations." M.I.T. Press, Cambridge, Massachusetts, 1966.

HUREWICZ, W., AND WALLMAN, H.
[1] "Dimension Theory." Princeton Univ. Press, Princeton, New Jersey, 1948.

KAPLANSKY, I.
[1] "Set Theory and Metric Spaces." Allyn and Bacon, Boston, Massachusetts. 1972.

KELLEY, J. L.
[1] "General Topology." Van Nostrand-Rheinhold, Princeton, 1955.

KERVAIRE, M. A.
[1] A manifold which does not admit any differentiable structure, *Comm. Math. Helv.* **34** (1960), 304–312.

KERVAIRE, M. A., and MILNOR, J. W.
[1] Groups of homotopy spheres: I, *Annals of Math.* **77** (1963), 504–537.

KLEIN, F.
[1] "Gesammelte Abhandlungen," Vol. I. Springer, Berlin, 1971.

KLINGENBERG, W.
[1] "Riemannian Geometry," Warter de Gruyter, Berlin, 1982.

KOBAYASHI, S., AND NOMIZU, K.
[1] "The Foundations of Differential Geometry," Vols. I and II. Wiley (Interscience), New York, 1963.

KOSZUL, J. L.
[1] Homologie et cohomologie des algèbres de Lie, Bull. Soc. Math. France. 78 (1950), 65–127. (MR 12, 120.).

LANG, S.
[1] "Differential Manifolds." Addison-Wesley, Reading, Massachusetts, 1972. (Supercedes "Introduction to Differentiable Manifolds." (Interscience), New York, 1962.)

LAWSON, H. B.
[1] "The Quantitative Theory of Foliations," *Reg. Conf. Ser. Math. No.* 27, American Mathematics Society, Providence, Rhode Island (1977).

LEHNER, J.
[1] "Discontinuous Groups and Automorphic Functions." Amer. Math. Soc., Providence, Rhode Island, 1964.

LEVI-CIVITA, T.
[1] Nozione di parallelismo in una varieta qualunque e consequente specificazione geometrica della curvature Riemanniana, *Rend. Palermo* **42** (1917), 73–205.

MASSEY, W. S.
[1] "Algebraic Topology: An Introduction." Harcourt, New York, 1967.

MILLMAN, R. S., and PARKER, G. D.
[1] "Elements of Differential Geometry." Prentice Hall, Englewood Cliffs, New Jersey, 1977.

MILNOR, J. W.
[1] "Morse Theory," Ann. of Math. Studies, Vol. 51. Princeton Univ. Press, Princeton, New Jersey, 1963.
[2] "Topology from the Differentiable Viewpoint." Univ. of Virginia Press, Charlottesville, Virginia, 1965.
[3] A unique decomposition theorem for 3-manifolds, *Amer. J. Math.* **84** (1962), 1–7.
[4] On manifolds homeomorphic to the 7-sphere, *Ann. of Math.* (2) **64** (1956), 399–405.
[5] Differential Topology. *In* "Lectures on Modern Mathematics," Vol. II (T. L. Saaty, ed.). Wiley, New York, 1964.
[6] Analytic proofs of the "hairy ball theorem" and the Brouwer fixed point theorem, *Amer. Math. Monthly* **85** (1978), 521-54.

MUNKRES, J. R.
[1] "Elementary Differential Topology." Ann. of Math. Studies, Vol. 54, Princeton Univ. Press, Princeton, New Jersey, 1963.
[2] "Topology, A First Course." Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

MYERS, S. B., AND STEENROD, N. E.
[1] The group of isometries of a Riemannian manifold, *Ann. of Math.* **40** (1939), 400–416.

NOMIZU, K.
[1] "Lie Groups and Differential Geometry." *Publ. Math. Soc. of Japan*, No. 2, Tokyo, 1956.

O'NEILL, B.
[1] "Elementary Differential Geometry." Academic Press, New York, 1966.

RIEMANN, B.
[1] Über die Hypothesen, welche der Geometrie zugrunde liegen, Ges. Werke, Leipzig, 1892. (Reprinted by Dover, New York, 1953.)

SAMELSON, H.
[1] Topology of Lie groups, *Bull. Amer. Math. Soc.* **58** (1952), 2–37.

SCHUR, F.
[1] Uber den Zusammenhang den Räume konstanter Krummungs-masses mit den projektiven Raumen, *Math. Ann.* **27** (1886), 537–567.

SCHWEITZER, P. A.
[1] Counterexamples to the Seifert conjecture and opening of closed leaves of foliations. To appear. *Ann. of Math.*

SEIFERT, H., and THRELFALL, W.
[1] "Lehrbuch der Topologie." Chelsea, New York, 1947.

SIEGEL, C. L.
[1] "Topics in Complex Function Theory," Vols. 1–3. Wiley (Interscience), New York, 1969–1973.

SINGER, I. M., AND THORPE, J. A.
[1] "Lecture Notes on Elementary Topology and Geometry." Scott-Foresman, Glenview, Illinois, 1967.

SMALE, S.
[1] Generalized Poincaré's conjecture in higher dimensions, *Bull. Amer. Math. Soc.* **60** (1960), 373–375.
[2] Dynamical systems, *Bull. Amer. Math. Soc.* **73** (1967), 747–817.

SPIVAK, M.
[1] "Calculus on Manifolds." Benjamin, New York.
[2] "A Comprehensive Introduction to Differential Geometry," 5 vols., 2nd ed. Publish or Perish, Inc., Wilmington, Delaware, 1979.

SPRINGER, G.
[1] "Introduction to Riemann Surfaces." Addison-Wesley, Reading, Massachusetts, 1957.

STEENROD, N. E.
[1] "The Topology of Fibre Bundles." Princeton Univ. Press, Princeton, New Jersey, 1951.

STERNBERG, S.
[1] "Lectures on Differential Geometry." Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

STOKER, J. J.
[1] "Differential Geometry." Wiley (Interscience), New York, 1969.

SUSSMAN, H. J.
[1] Orbits of families of vector fields and integrability of distributions, *Trans. Am. Math. Soc.* **180** (1973), 171–188.

WALLACE, A.
[1] "Differential Topology, First Steps." Benjamin, New York, 1968.

WARNER, F.

[1] "Differentiable Manifolds and Lie Groups." Scott-Foresman, Glencoe, Illinois, 1972.

WEYL, H.

[1] "Die Idee der Riemannschen Fläche." Teubner, Leipzig, 1923. [*Engl. transl.*: "The Concept of a Riemann Surface." Addison-Wesley, Reading, Massachusetts.]

[2] "Selecta." Birkhäuser, Basil, 1956.

WHITEHEAD, J. H. C.

[1] Convex regions in the geometry of paths, *Quart. J. Math.* **3** (1932), 33–42, 226–227.

WHITNEY, H.

[1] Differentiable manifolds, *Ann. of Math.* (2) **37** (1936), 645–680.

WOLF, J. A.

[1] "Spaces of Constant Curvature," 2nd ed. Publish or Perish, Inc., Wilmington, Delaware, 1978.

ZASSENHAUS, H.

[1] "The Theory of Groups." Chelsea, New York, 1958.

INDEX

E

F

N

O

P

Q

R

S

T

U

V

W

X

Z